Neurofanos

Internship report

Neurofanos

School tutor : CONZE Pierre-Henri

Company tutor : ABBASI Hamid

Student: BOULEFRED-DERRAR Inès

April 28th 2025 - October 24th 2025

Multi-class neuronal cell classification in histological images from HI-impacted fetal sheep using deep neural networks

I. Acknowledgments

I would like to thank my tutor Dr. Hamid Abbasi, for this opportunity as well as his goodwill and kindness during my internship.
I would like to thank Callan Loomes for his valuable insights and answers to my numerous questions during this time, as well as Nima Sadeghzadeh for his kindness, support and his immediate willingness to integrate me into the Auckland Bio-engineering Institute and NeuroTech Lab community.

II. Summary

Hypoxic-ischemic encephalopathy (HIE) refers to a condition where the brain receives inadequate oxygen (hypoxia) due to reduced or blocked blood flow (ischemia) around the time of birth. This may happen in cases such as when newborn babies are born with umbilical cords coiled around their neck. This combination disrupts both oxygen delivery and the brain's ability to clear away waste products like carbon dioxide and other metabolic byproducts, leading to cellular damage which can lead to lifelong neurodevelopmental complications. Therapeutic hypothermia is currently the only recognised treatment for the condition. However, it must be applied within the first 6 hours following ischemia for it to be effective. Drug development is a promising alternative to make access to care much easier, faster and less risky, but it is hindered by the necessity of manual cell counting in histological images to quantify levels of damage.
This project aims to automate and accelerate this step using deep learning models for cell classification in fetal sheep brain histology affected by HIE to varying degrees. A hybrid version of a Vision Transformer (ViT)-CNN, a custom made CNN model, two different ViTs pre-trained on different image sizes (384×384 and 224×224), Swin, SwinV2, EfficientNet, ResNet50, and a ViT pre-trained on histological images were finetuned using a dataset of 4263 images to be classified into three classes: Normal, Intermediate and Pyknotic, depending on the extent of cell damage. The dataset was manually segmented classified and was composed of histological images from six different brain regions of the fetal sheep (CA1, CA3, CA4, DG hippocampal regions and PS1, PS2 parasagittal cortical regions) across three treatment groups (sham, hypoxia treated with hypothermia and hypoxia without treatment). Results so far show that the models have performed with very close validation accuracies, with the top performing models being SwinV2 (82.01%), the hybrid ViT-CNN model (81.80%) and the custom CNN model (81.46%). They show that varying architectures have very similar performance results, indicating a performance saturation on the given training dataset, with areas under the curve of the micro-averaged receiver operating characteristic (ROC) curve being 0.93, 0.90 and 0.93 for the SwinV2, hybrid ViT-CNN and custom CNN models respectively.

This work is useful for extending datasets that require expert inputs and are laborious and time-consuming to expand manually, as well as for classifying cells to aid in the development of preclinical drugs for treating hypoxic-ischemic encephalopathy.

III. Table of contents

IV. Internship location

A. Auckland Bioengineering Institute

The Auckland Bioengineering Institute (ABI) is a research center within the University of Auckland that was founded in 2001. Its work is multidisciplinary, spanning : engineering, medicine, computer science, and the life sciences.

The ABI's mission is to apply engineering principles and computational approaches to the study of biological systems.

The institute is structured around a high amount of research domains in the field of biology : animals and microbiomes, bioinstrumentation, biomimetics, biorobotics, cancer imaging, cardiac electrophysiology, circulation and transport, computational methods, empathic computing, eye health and diagnostics, the gastro-intestinal system, heart mechanics, implantable devices, computational methods, animate technologies, liver modelling, the lungs and respiratory system, the musculoskeletal system, the nervous system, reproductive systems, sports biomechanics, women's health, surgical engineering and lastly medical imaging and biomedical informatics.

The institute has a strong entrepreneurial impact: over the past decade, more than twenty spin-out companies have been created from ABI research.

Alongside these activities, the ABI has developed its own programs for engaging with industry and fostering innovation. The Cloud 9 collaboration initiative is one such program, designed to strengthen connections between ABI researchers and the business community.

My internship took place at Neurofanos, a start-up of the NeuroTech lab, a laboratory for computer science research in neurology at the ABI.

B. The NeuroTech Lab

The NeuroTech Lab is a multidisciplinary research group operating within the University of Auckland, primarily through the Centre for Brain Research and the Auckland Bioengineering Institute, in collaboration with the Department of Neurosurgery at Auckland City Hospital. It is led by Dr. Hamid Abbasi. Its goal is to transform neurosurgery and neuroscience through technology using artificial intelligence, advanced imaging, and other engineering technologies at the service of surgical practice and the understanding of brain disease.

The lab collaborates with many partners (cf. next section), drawing on expertise from neurosurgeons, engineers, AI specialists, pathologists, physiologists, and other scientists.

Collaborations extend to the Matai Medical Imaging Institute, the Department of Physiology, and the Centre for Cancer Research.

In terms of current work, Research fellow Sam Cutfield has led the development of real-time molecular fingerprinting of brain tumours using Raman spectroscopy in order to provide instant intraoperative diagnosis information.

Postdoctoral fellows Maryam Tayebi and Alireza Sharifzadeh-Kermani are developing what they describe as a “live brain GPS,” combining diffusion MRI tractography with biomechanical modelling to compensate for brain shift (the deformation and displacement of brain tissue during surgery). The aim is to develop patient-specific brain maps that can be updated continuously in real time in the future.

Similarly, PhD candidate Ali Roozbehi is creating pipelines for brain-shift prediction, integrating multimodal imaging with AI to track cortical movement.

Jiantao Shen is working on automated 3D brain reconstruction and meningioma segmentation.

Nima Sadeghzadeh has developed AI methods to predict meningioma growth from a single MRI scan.

Callan Loomes is applying AI to automate glial segmentation, enabling large-scale mapping of neuroinflammation, while also creating 3D-printed brain models to support surgical training.

Naima Noor is linking MRI imaging with epigenomic data to accelerate meningioma diagnosis, a project supported by both the Centre for Brain Research and the Centre for Cancer Research.

Sanaz Movahed is working on AI-driven tumour subtyping using DNA methylation data and digital pathology, in order to provide classification to support intraoperative and diagnostic decisions.

Shanan Chand has applied AI techniques to chemical space exploration, developing pipelines to accelerate drug discovery and collaborating with pharmacologists to uncover new therapeutic molecules.

Manpreet Kaur has designed an AI-based pipeline for motion tracking of infants’ general movements, in order to detect neurological risk factors such as cerebral palsy. This has been developed into a mobile application, TinyMotion.

Another PhD project by Jiangfan Yu targets brain MRI motion artifacts, creating deep-learning methods to denoise said artefacts and thereby enhance the reliability of motion-sensitive biomarkers.

Several members of the NeuroTech lab such as Callan Loomes, Ali Roozbehi, Maryam Tayebi and Alireza Sharifzadeh-Kermani and Jiantao Shen and their work are directly involved in Neurofanos, a start-up company developing AI-driven intraoperative neuronavigation.

The lab is inherently multi-cultural with its people being from various countries in Asia, Europe, and New-Zealand.

I have had the opportunity to work alongside most of these people during my internship.

C. Neurofanos: AI-Driven Intraoperative Neuronavigation

Neurofanos is a Kiwi medtech startup established in 2023, based in Auckland. Its aim is to revolutionize neurosurgery through artificial intelligence-powered intraoperative neuronavigation. The company aims to achieve this by equipping neurosurgeons with real-time, patient-specific guidance that enhances precision and safety during brain surgery. Their technology aims to integrate real-time visual insights with scalable design, supporting both preoperative planning and intraoperative decision-making to reduce surgical complications and improve outcomes.

The company first gained attention after winning the University of Auckland's Velocity $100k Challenge in 2022, which provided them with seed funding and access to VentureLab's incubator program. Since then, they've continued to build momentum and were recently awarded NZD 1 million through the Ministry of Business, Innovation and Employment's Endeavour 'Smart Ideas' Fund. The company is currently further developing their neuro-navigation technology in partnership with the Auckland Bioengineering Institute and has developed a proof of concept prototype that has been endorsed by neurosurgeons.

Today, Neurofanos' strategy is to continue developing key strategic partnerships with industry actors. For example, Neurofanos is now part of the NVIDIA Inception Program which offers startups access to developer resources, preferable pricing on NVIDIA products and exposure to the venture capital community among others.

After the initial proof of concept, a key challenge for Neurofanos is to continue building up the necessary foundation for product development, through continuous research efforts.

a. Organization and people

Neurofanos is what can be considered as a university spin-out company, and its structure is what can be described as an "adhocracy", meaning that it has a fluid and flexible management type, at least at the lower levels. The research work is not organized around specific job titles but rather around specific missions, like in a research team. Access to

broader information concerning product development is subject to intellectual property and not readily available to the general public, including to me as an intern.

Founders and leaders include Dr Hamid Abbasi, an AI and machine-learning specialist, Dr Jason Correia, a consultant neurosurgeon, and Dr Samantha Holdsworth, a medical imaging scientist. The broader team encompasses PhD and post grad algorithm developers and specialists such as Jiantao Shen, Callan Loomes, Alireza Sharifzadeh, and Ali Roozbehi, supported by imaging and software engineers like Dr Maryam Tayebi and Dr Mahyar Osanlouy, all members of the NeuroTech lab at the Auckland Bioengineering Institute.

The company also has advisors in various fields (investment, neuropharmacology and imaging).

b. My place in the team

During my internship, I worked autonomously on a project that was a part of PhD candidate Callan Loomes's broader research, the classification of neuronal cells from HIE (Hypoxic-ischemic encephalopathy)-affected fetal sheep models developed by the Department of Physiology. Callan's work aims to establish that automated cell analysis could replace manual counting and have immediate pre-clinical impact on the development of new treatment strategies for HIE affected births in humans. The next step was testing whether transformer models could improve on his already established classification system, particularly for distinguishing subtle differences between intermediate and severely damaged neurons, which is the mission that was given to me.

The impact of this mission is to provide additional information on the ability of deep learning models to provide an automatic pipeline for cell identification and counting in the brain.

Callan's desk was just steps away from mine in our open-plan workspace, as were the desks of most other members of the NeuroTech lab and Neurofanos team.

Although I didn't collaborate regularly with other NeuroTech Lab members in terms of technical work, their proximity made seeking advice, help and feedback straightforward and easy.

My internship offered me a high degree of independence and flexibility. I managed my daily tasks from my desk in the open space, collaborating closely with my company tutor Mr. Abbasi, whose office was quite close to my desk. This proximity meant I always had access to support from my tutor or Callan when needed, and had the freedom to work at my own pace.

Nearly every week I had the opportunity to attend different presentations or seminars, and social events were scheduled weekly as well.

D. Partnerships relevant to my internship

The Department of Physiology at the University of Auckland provided the experimental foundation for the work by providing histological images of HIE-affected brain cells in fetal sheep and thus the foundation upon which the datasets were built. Professor Alistair Gunn, Professor Laura Bennet, Dr Guido Wassink, and Dr Joanne Davidson are part of their academic staff. Their team has developed fetal sheep models that replicate brain injury in newborns, allowing them to control when injury occurs and monitor the resulting brain changes over time (cf. Mission - Dataset construction)

Their work focuses on therapeutic hypothermia, a treatment where newborns are cooled after birth complications to reduce brain damage.

The NeuroTech Lab focuses on developing machine learning algorithms that analyze the brain signal data collected from the physiology experiments. These algorithms are designed to automatically detect seizures and early signs of brain injury that might be missed by human observers.

Working with researchers S. Dhillon and K. Zhou, Dr. Abbasi focuses on making these detection systems work in real-time hospital environments. The team uses advanced signal processing techniques and neural networks to identify patterns in brain activity that predict injury development. The goal is to provide early warnings that could trigger protective treatments before permanent damage occurs.

Starship Children's Hospital contributes human patient data and clinical guidance to the research.

The Mātai Medical Imaging Institute supplies brain imaging data that serves as validation for the detection algorithms.

V. Mission context

A. Hypoxic Ischemic encephalopathy

Hypoxic-ischemic encephalopathy (HIE) is one of the reasons for morbidity and mortality of newborns all over the world [11]. It is caused by a hypoxic-ischemic (HI) event, a brain injury due to a lack of oxygen and blood flow to the brain, during the prenatal, intrapartum or postnatal period, in which it can develop acutely or chronically [4,13]. Perinatal HIE occurs in 1–3 per 1000 live births [1] in developed countries and 10–20 per 1000 live births in low and middle-income countries [10]. Among these cases, 15–20% of the affected infants may not survive beyond the age of 2[2]. Another 25% of the affected newborns will be subject to life-long complications, such mental retardation, visual motor or visual perceptive dysfunction, auditory impairment, increased hyperactivity, cerebral palsy, and epilepsy [3].

HIE is an evolving process during which cell death occurs gradually, in which a primary phase, a latent phase, a secondary phase and a tertiary phase occur sequentially between the HI event and the final outcome, the brain injury [8].
There are currently several treatments available to newborns affected by HIE, the most common of which being mild induced hypothermia, which can effectively reduce the risks of death and disability after a moderate to severe HI event [6].
Hypothermia treatment is most effective when administered within a certain window of opportunity [7]. Studies have shown that starting hypothermia during the latent phase of HIE (i.e., the first three hours after ischemia ends) can lead to recovery of brain activity close to baseline levels [7]. However, as is often the case, the HI event may occur well before birth, [7] making the delay between injury and treatment unknown. A controlled randomized study showed that hypothermia could only be performed on 12% of affected infants within a four–hour window of birth [9]. Mild hypothermia, as a standard of care, has also been paired with other neuroprotective agents, the most promising of which being stem cells and rhEPO (recombinant human erythropoietin) [8].

Moreover, hypothermia, while beneficial, is not a cure. Even when delivered optimally, nearly half of cooled infants still die or develop disability [16]. This has motivated an intensive search for alternative therapies, including pharmacological agents such as recombinant human erythropoietin (rhEPO), xenon gas, stem cell therapies, and antioxidants. The Department of Physiology at the University of Auckland has been central to these efforts, combining preclinical work in fetal sheep with translational neuroprotection trials. Yet, drug discovery and validation remain hindered by methodological bottlenecks. A key challenge is the quantification of neuronal injury in experimental histology: assessing cell death, morphology, and regional vulnerability currently requires painstaking manual counting by expert neuropathologists. This process is slow and subjective [12].

Animal models have been essential to building this knowledge. EEG, MRI, and histological assessments from these models have revealed the early biomarkers of injury, the optimal

therapeutic windows, and the cellular correlates of outcome. The fetal sheep model is a well-established pre-clinical standard for studying human hypoxic-ischemic encephalopathy (HIE), as the fetal sheep brain is gyrencephalic and similar in maturity to a term human infant. This data is useful for both understanding injury mechanisms and for training and validating computational tools such as AI models.

The clinical urgency of HIE, the limitations of current therapy, and the laborious nature of histological assessment define the context for the work that I took a part in. Automated neuronal cell classification, using state-of-the-art deep learning models such convolutional neural networks, vision transformers and hybrids, represents a possibility to accelerate preclinical therapy development.

B. Progress and objectives

The wider initiative to which this project belongs is designed to contribute to the early identification and diagnosis of hypoxic–ischemic encephalopathy (HIE) in newborns. The objectives include :

a. The determination of ‘timing biomarkers’, electrophysiological biomarkers of injury timing and severity based on EEG data.
b. The development of machine learning algorithms in order to predict the severity and the phases of injuries.

a. EEG biomarkers and real-time diagnosis

The first component of the project is concerned with the development of deep-learning algorithms to identify timing biomarkers in the hours immediately following a hypoxic–ischemic event. The core hypothesis, supported by experimental work in fetal sheep, is that the temporal evolution of the EEG signal contains critical signatures of injury progression and therapeutic windows. Studies by Gunn, Bennet, Davidson and colleagues have demonstrated that after an HI injury, the EEG passes through characteristic phases that represent latent, secondary, and tertiary injury cascades.

Experimental studies in fetal and neonatal sheep have provided a platform for decoding these signatures, through continuous high-resolution electrophysiological recording and precise control of injury timing. Recent advances [17, 18, 19, 20] have further refined these markers by linking time–frequency features and seizure burden to long-term outcomes. The integration of wavelet analyses, grey-level co-occurrence methods, and time–frequency coherence has created a robust feature space for automated algorithms.

The next stage is to embed these feature sets into deep neural networks capable of online classification and prediction. Early work has shown the feasibility of convolutional and recurrent architectures for seizure detection and timing prediction [21], while ongoing efforts now test transformer-based models for temporal sequence learning. The final objective is to predict the severity and phase of brain injury in real time, directly from EEG streams, thereby guiding interventions such as therapeutic hypothermia during the narrow window in which they remain effective.

b. Prediction of the severity and phases of injury

Complementing the macroscopic EEG work, the second component focuses on microscopic histological analysis. Quantification of neuronal survival and death in brain regions affected by HIE has relied on labor-intensive manual counting of neurons across histological images. This bottleneck slows preclinical therapeutic development. To address this, Callan Loomes developed a two-step automated segmentation–classification pipeline [12].

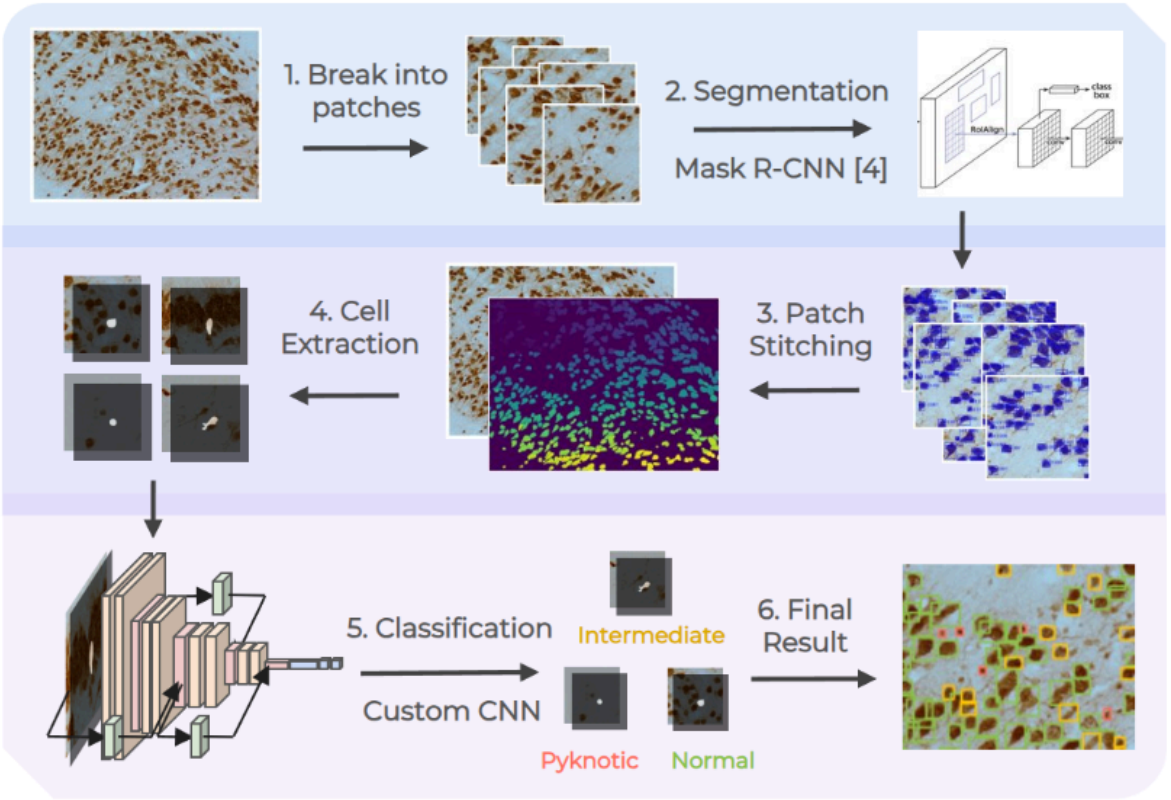


figure 1 : Established classification and segmentation pipeline

In this approach, histological images from fetal sheep subjected to sham, hypoxia+hypothermia, and untreated hypoxia were segmented using a generalized Mask

R-CNN trained on thousands of manually annotated neuron instances, achieving average precision of ≈88% at IoU 0.5. Segmented neurons were then classified into three morphological states (Normal, Intermediate, and Pyknotic) using a lightweight custom CNN trained on an annotated dataset. The classifier achieved ≈93% accuracy, and the pipeline's automated counts correlated strongly with the manual counts across experimental groups and brain regions. This confirmed that automated histology can produce reliable information on injury severity, suitable for translational and therapeutic studies.

C. My Mission

As said before, the main problem this branch of the global project means to tackle is the time and resource consuming nature of the cell identification and counting process. In order to bring a solution to this problem, Callan's work focused on developing a segmentation and classification pipeline using the available ground truth of 1500 cells. The complete dataset, including this ground truth, comprises 44000 cells in total.

My mission is to directly extend Callan's baseline CNN proposed solution by exploring the performance of transformer models noted in recent literature as state of the art. My internship mission was to establish the most efficient neural network model among a list of chosen and fine tuned models (including Vision Transformers, Convolutional Neural Networks and hybrids between the two) for the classification of HIE affected neuronal cells from six different brain regions of the fetal sheep models (provided by the Department of physiology) into three morphological categories : healthy, intermediate and pyknotic (respectively increasingly affected by ischemia).

Models were assessed using accuracy, metrics deducted from confusion matrices, ROC curves, inference time, and parameter efficiency, in order to identify the most effective architecture for integration into the segmentation–classification system.

This work serves the following purposes :

1. Extending the existing dataset of 1500 labeled histological images of cells across six different brain regions of fetal sheep brains. This will make the development of increasingly accurate and powerful segmentation and classification models possible.
2. Establishing the most efficient and accurate classification model to-date for the segmentation and classification pipeline, thereby improving its automatic neuron counting and classification capabilities.

D. Planning and management

A detailed Gantt chart (cf. Annex 1) was developed and presented during the second project meeting with the project team at the ABI, namely Dr. Hamid Abbasi and Callan Loomes. This chart served as the primary tool for planning, tracking, and visualizing the project timeline and major milestones from May to October 2025.

The Gantt chart provides an overview of the project's three main phases:

- Project Familiarization (April - May 2025): This initial phase was dedicated to getting up to date in terms of technical knowledge. This included reviewing research papers on ischemia-hypoxia, getting informed on the study's subject, studying the previous work on classification and segmentation, learning about transformer architecture and extending the dataset manually in order to have additional data for training and to get familiar with the dataset. This was done by segmenting 36 images using the Napari software. It also included a small period of training, through a machine learning workshop.
- Comparative Study (May - October 2025): This is the main phase of the internship. It began with reviewing literature and the Hugging Face resource in order to select the models for the comparative study. Then, the code for training was built, and training as well as the evaluation of the models ensued. The coding process involved data augmentation, modifying each model's architecture to ensure the correct input and output requirements were met, hyperparameter setting, K-fold cross-validation setup, and setting up evaluation using metrics like confusion matrices, ROC curves, and resource usage measurements.
- Manuscript Writing (June - October 2025): This phase involves the writing of both the internship report and an academic journal paper to document the work.

I had a high degree of autonomy in my daily work, and managed my schedule using this plan. Project management and technical guidance were provided by Dr. Abbasi and Callan Loomes. I reported my progress to them regularly through meetings or informal exchanges, which varied between weekly, bi-weekly, and monthly reviews based on the project advancement.

E. Socio-environmental challenges

1. The Social Contribution of HIE Research

Hypoxic-Ischemic Encephalopathy (HIE) is a leading cause of mortality and long-term neurological disabilities in newborns, including cerebral palsy, epilepsy, and cognitive deficits. Its impact on society and healthcare infrastructures is considerable. This internship contributes to alleviating that burden. By accelerating pre-clinical evaluation, this approach shortens the time from potential therapy discovery to clinical application, supporting the

development of drugs for HIE and ultimately improving survival and quality of life for affected infants.

2. Ethics regarding the dataset and the use of AI

One of the most important issues regarding the use of AI is its propensity towards bias through the biases of the datasets used. An advantage of this project is that the data comes exclusively from animal models, which considerably if not completely reduces the chances of the ensuing segmentation-classification pipeline having any tendency towards discrimination between humans.

The histological image data used for model training in this project was obtained from a pre-existing, ethically approved animal study.

The detailed methodology for obtaining the brain tissue and subsequent histology can raise ethical questions, as this data stems from experiments performed on animals. All procedures were approved by the animal ethics committee of the University of Auckland under the New Zealand Animal Welfare Act, and the code of ethical conduct for animals in research established by the ministry of primary industries of the government of New Zealand.

Romney/Suffold fetal sheep at 118 to 124 days of gestation were subject to experiments. Food but not water was withdrawn 18 hours before surgery. The ewes were under anesthesia during surgery, and the depth of anesthesia, maternal heart rate and respiration were constantly monitored. The ewe's abdominal midline was incised, and the fetus was exposed and both fetal brachial arteries were catheterized to measure blood pressure. Other vitality measures were carried out using catheters and electrodes.

Inflatable carotid occludes were placed around both carotid arteries in order to induce ischemia with sterile saline for 30 minutes. A cooling cap was then placed on the fetal head. The uterus of the ewe was then closed and all measuring instruments were removed. The fetus and ewes then underwent post-operative care in a controlled environment. [6]

3. Environmental Impact of AI and the NeSI Infrastructure

The training and fine-tuning of large AI models are computationally intensive processes with a significant carbon footprint. This environmental impact stems from the electricity consumption of high-performance computing (HPC) clusters powered by GPUs, which in turn generates CO_2 emissions. The training and fine tuning of all models took place on the NeSI (New Zealand eScience Infrastructure) HPC and lasted for about 3 days.

An advantage of this project in terms of mitigating this impact is Transfer Learning. Instead of training models from scratch, models were already pre-trained on generic or specific image

datasets such as ImageNet. The process of fine-tuning requires significantly less computational time, and reduces the carbon footprint compared to training from scratch.

VI. Mission

A. Methodology

In order to successfully carry out the mission of comparing different model architectures on a dataset that was built using already available data and manually segmented data, the following procedure was carried out :

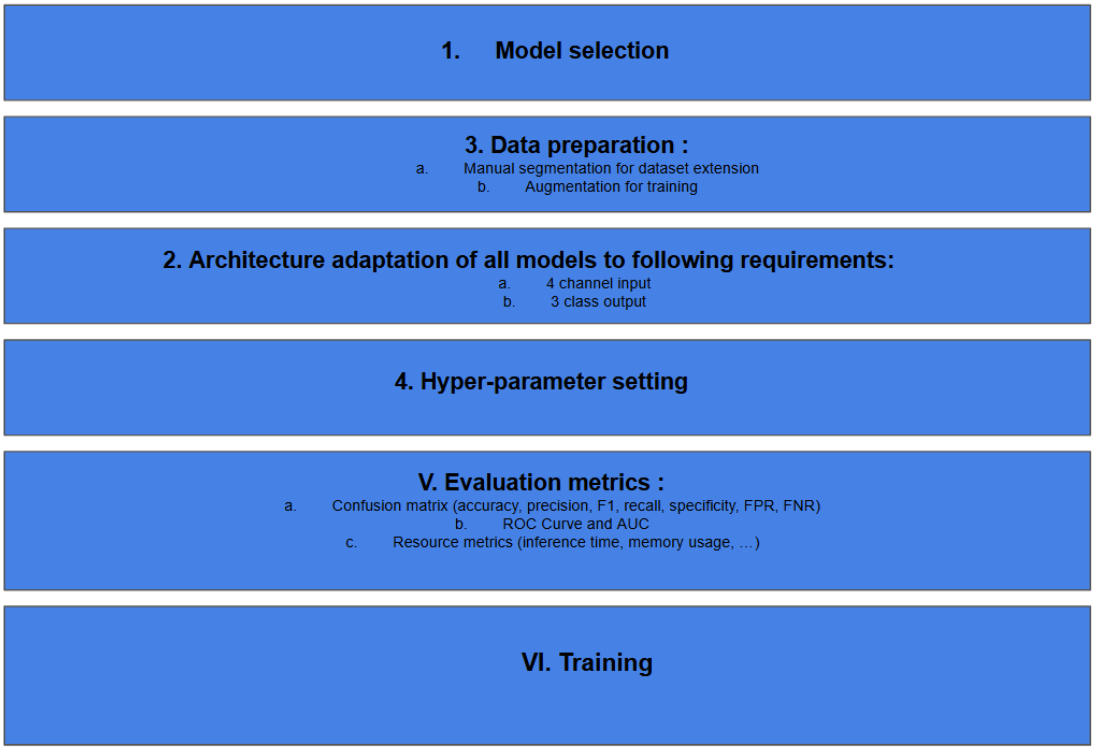


figure 2 : Outline of the study preparation

After training, the first results and metrics were obtained, analyzed and a more thorough training code was prepared in order to further the study and find additional information according to the interrogations the first results might have brought about.
Each step of the above-described methodology is described below.

a. Dataset construction

The data used in the project consists of high-resolution histological images of fetal sheep brain tissue, provided by the Department of Physiology of the University of Auckland (cf. Internship location; Partnerships).
The tissue samples were obtained from a controlled experimental study. Pregnant ewes were subjected to a procedure where the fetus underwent carotid artery occlusion to induce ischemic hypoxia (cf. socio-environmental challenges). The experimental groups were:

- Ischemia (I): Subjected to carotid artery occlusion.
- Ischemia + Hypothermia (I+H): Subjected to occlusion and treated with therapeutic hypothermia.
- SHAM: Control group that underwent a sham surgery without occlusion.

Examples of histological images from these three different groups can be found below :

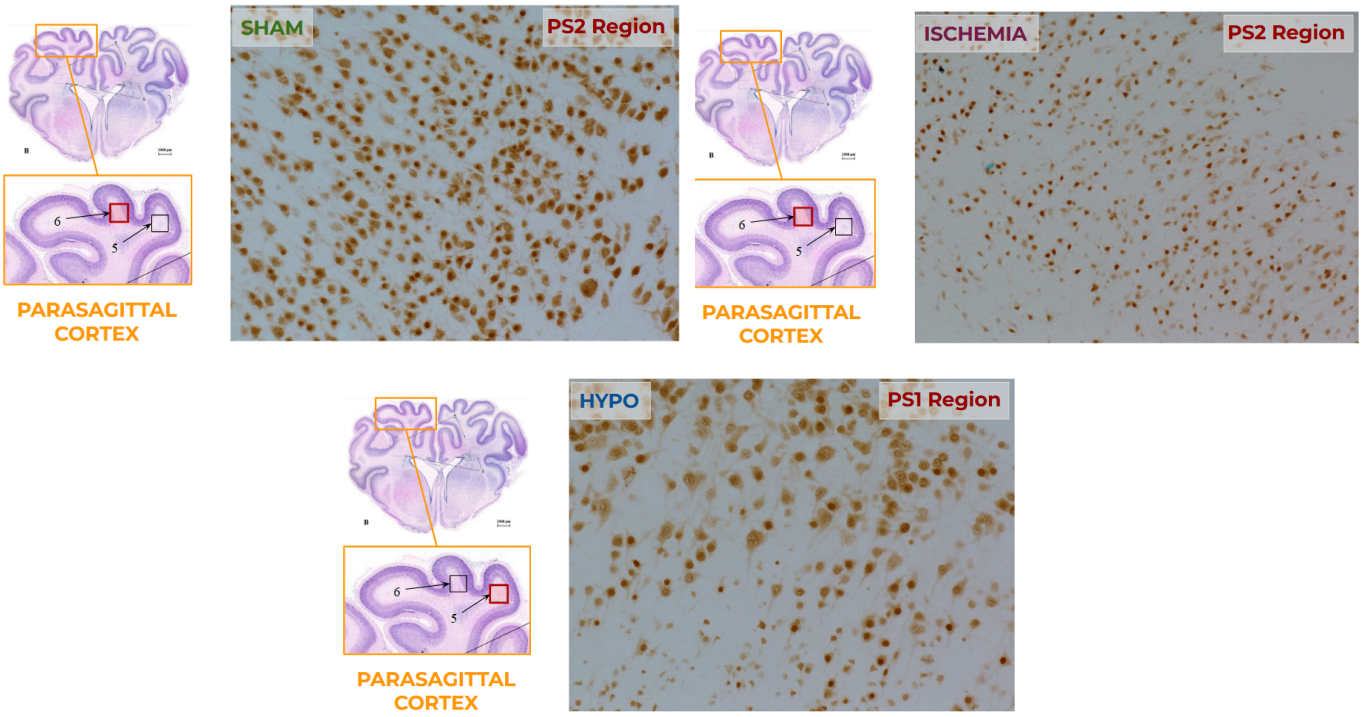


figure 3, 4, 5 : Histological images of the PS2 and PS1 region in the parasagittal cortex, showing cells from the sham, ischemia and ischemia+hypothermia groups

The analysis focuses on specific regions of the hippocampus and cortex. The six regions studied are: CA1, CA3, CA4, Dentate Gyrus (DG), Parasagittal Cortex 1 (PS1), and Parasagittal Cortex 2 (PS2).

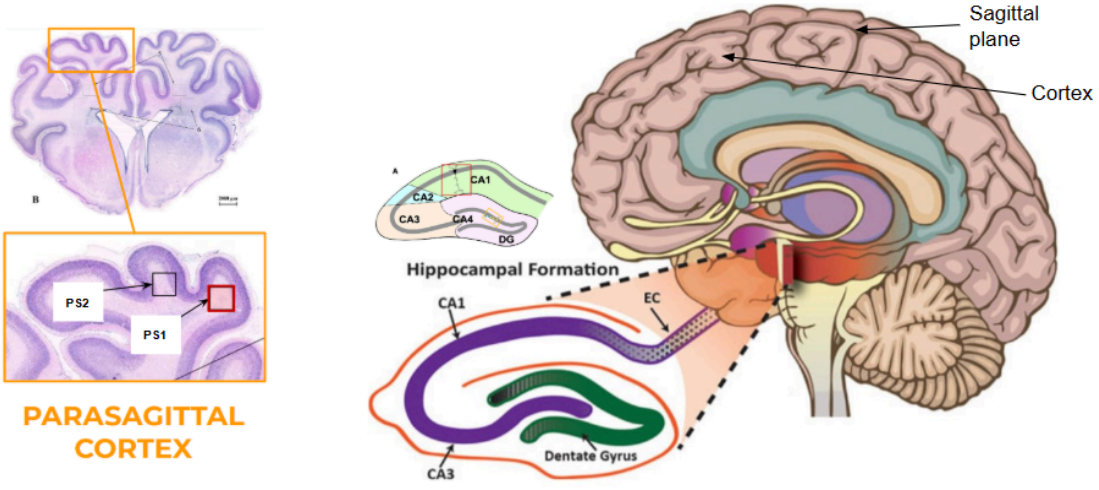


figure 6 : Visualization of the brain regions targeted by the study

The cells within these regions exhibit distinct morphological changes based on the severity of injury, which form the basis for their classification:

- Normal Cells: Uniform cells with clear nuclei, with an area above 2000, and a sphericity under 0.7.
- Intermediate Cells: Faint cells, where the nucleus fuses with the cytoplasm, with an area under 2000.
- Pyknotic Cells: Severely damaged, small, circular and darkly colored. They exhibit a white halo and have an area under 100 with a sphericity above 0.95.

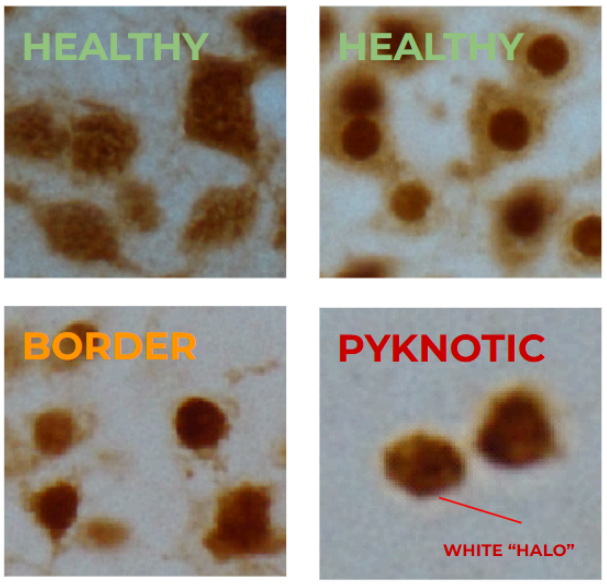


figure 7 : Morphology illustration of the normal (healthy), border (intermediate) and pyknotic cells

1,500 out of the available 44,000 cells delivered by the Department of Physiology were manually segmented and classified across the different brain regions and experimental groups for previous studies. This process involved precisely outlining each cell on the Napari software and assigning its class (Normal, Intermediate, Pyknotic) using a custom made interface. This initial dataset was used for the segmentation-classification pipeline that was already established.

One of the goals of my mission was to significantly expand this dataset using 36 additional histological images provided. These 36 additional images consisted of two histological images per brain region for each treatment group (example below).

I repeated the same process as was used for the construction of the initial dataset in order to segment and classify them.

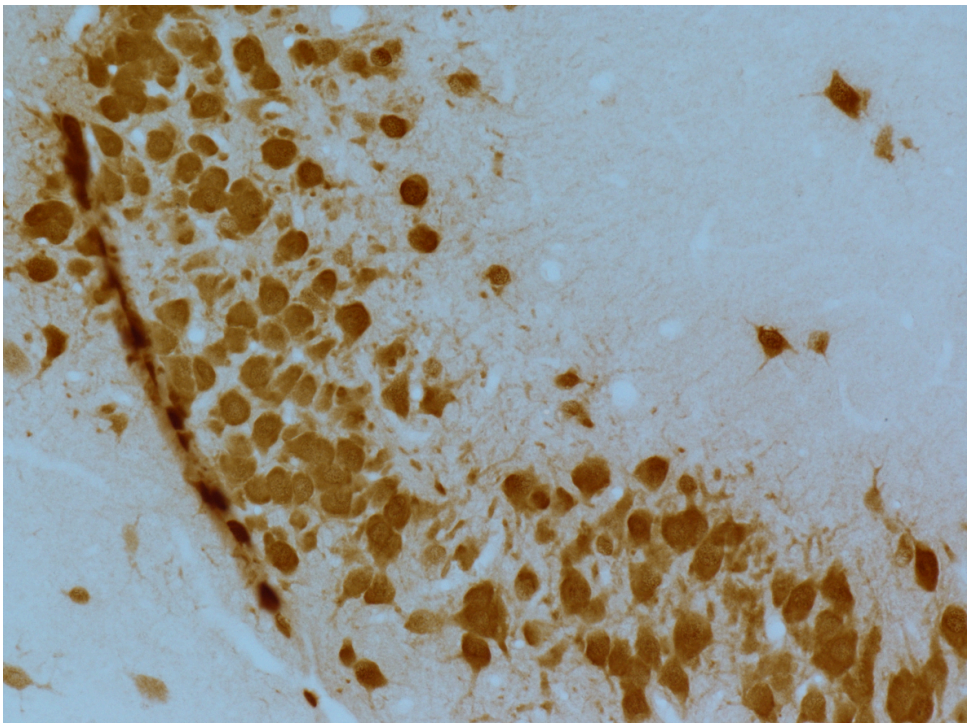


figure 8 : Example of a histological image of the CA3 region, from the ischemia+hypo group, to be manually segmented and classified.

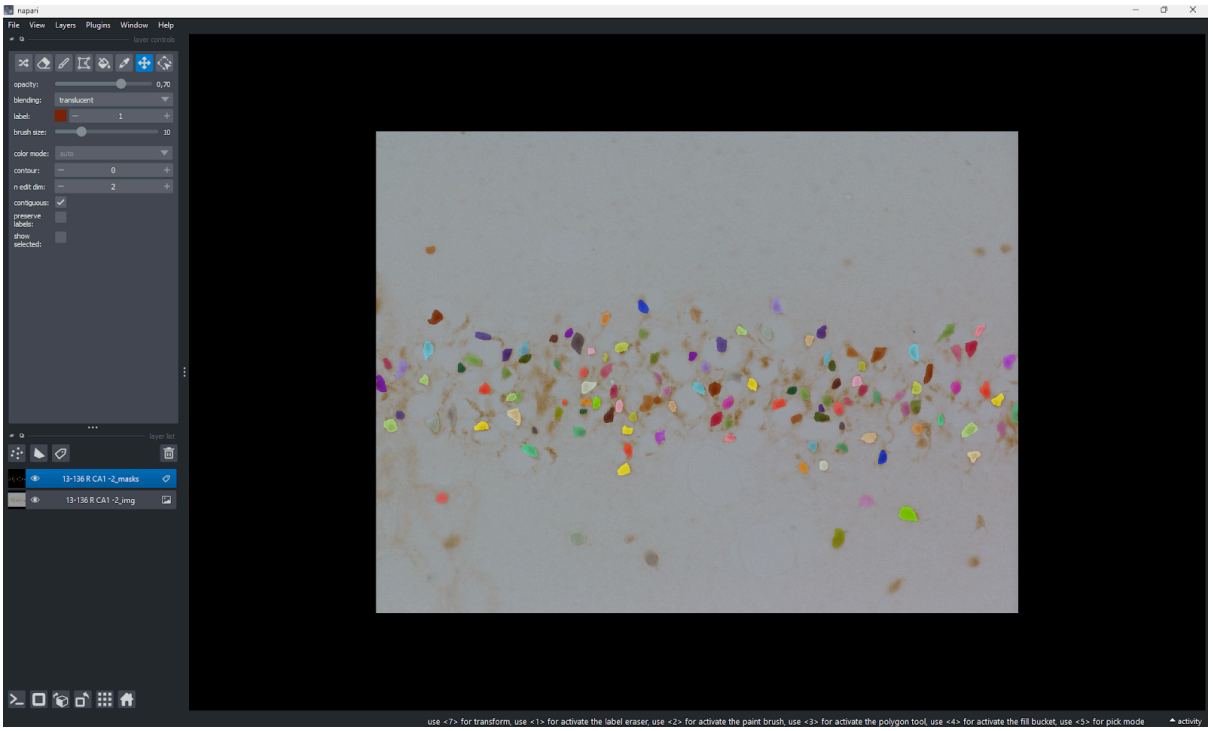


figure 9 : Napari interface for manual segmentation.

Final dataset : After manual segmentation and classification, the extension of the dataset included 4263 cells.

Number of pyknotic cells: 304

Number of normal cells: 1291

Number of intermediate cells: 2668

b. Model choice

The field of image classification has been revolutionized by deep learning. The current state-of-the-art is shifting from Convolutional Neural Networks (CNNs) to Transformer-based models, with hybrid approaches also holding promise.

For nearly a decade, CNNs were the state of the art benchmark. Architectures like ResNet and EfficientNet represent the state of the art. Their inductive bias (local feature extraction and translation invariance) makes them data-efficient and performant for computer vision tasks. They remain a strong baseline, especially when computational resources are limited.

Introduced in 2020 in the paper “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”, the Vision Transformer marked a shift. By adapting the Transformer architecture, originally designed for Natural Language Processing, ViTs treat an image as a sequence of patches. The core mechanism is self-attention, which allows the model to learn to weigh the importance of all other patches of an image when encoding a specific one. This grants ViTs a global vision from the first layer, unlike CNNs that build it gradually. Models like the ViT-Base/Large and Swin Transformer (which introduces a hierarchical shifted window mechanism to compute attention locally) have outperformed CNNs on large-scale benchmarks. Swin's hierarchical design allows it to capture features at the local and global scales, making it a powerful architecture.

Hybrid models combine convolutional layers with transformers. The ViT-Hybrid model used in this study uses a CNN backbone (BiT) as a feature extractor whose features are then used as initial tokens to create the patch embeddings for the transformer encoder. This leverages the CNN's strength in processing low-level, local features and the Transformer's strength in modeling long-range dependencies between these features.

State of the art for biomedical image analysis often involves models that are pre-trained on large-scale domain-specific datasets. For example, CellViT and PandaVit are Vision Transformers specifically pre-trained on histological image patches. This domain-specific pre-training allows it to achieve superior performance on tasks like cell segmentation and classification compared to models pre-trained on natural images (such as the ImageNet dataset).

While domain-specific models like CellViT represent the ultimate performance ceiling, different architectures trained and fine-tuned on the same datasets were an important characteristic to include in the model choice in order to deduce a model architecture comparison. This study benchmarks a suite of ImageNet-pretrained models (CNNs, ViTs, Hybrids).

Using ImageNet-pretrained models for almost all selected architectures provides a level playing field, and it eliminates the variables of different pre-training datasets. These controlled conditions will yield information regarding the architecture of the studied neural networks and their efficiency in the classification task.

For this study, eight distinct model architectures were implemented and compared, covering CNNs, pure Vision Transformers, hybrid models, and a custom CNN.

1. ResNet-50

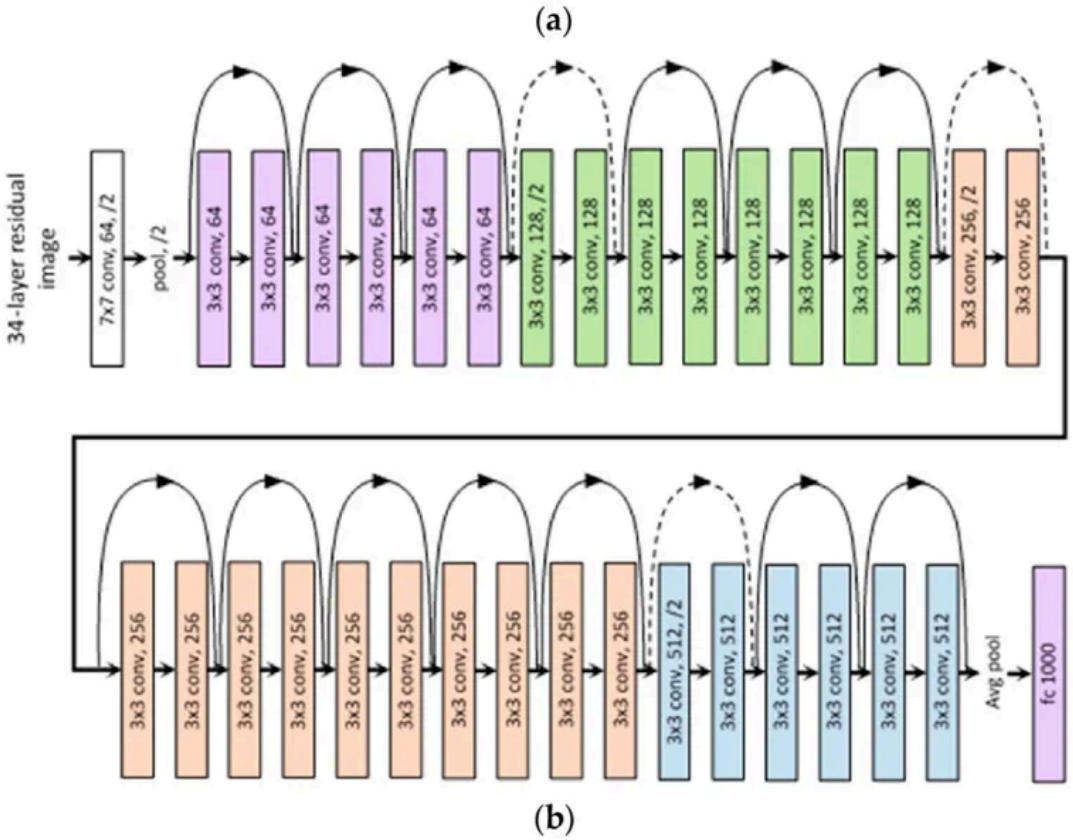


figure 10 : ResNet-50 architecture

2. EfficientNet-b7

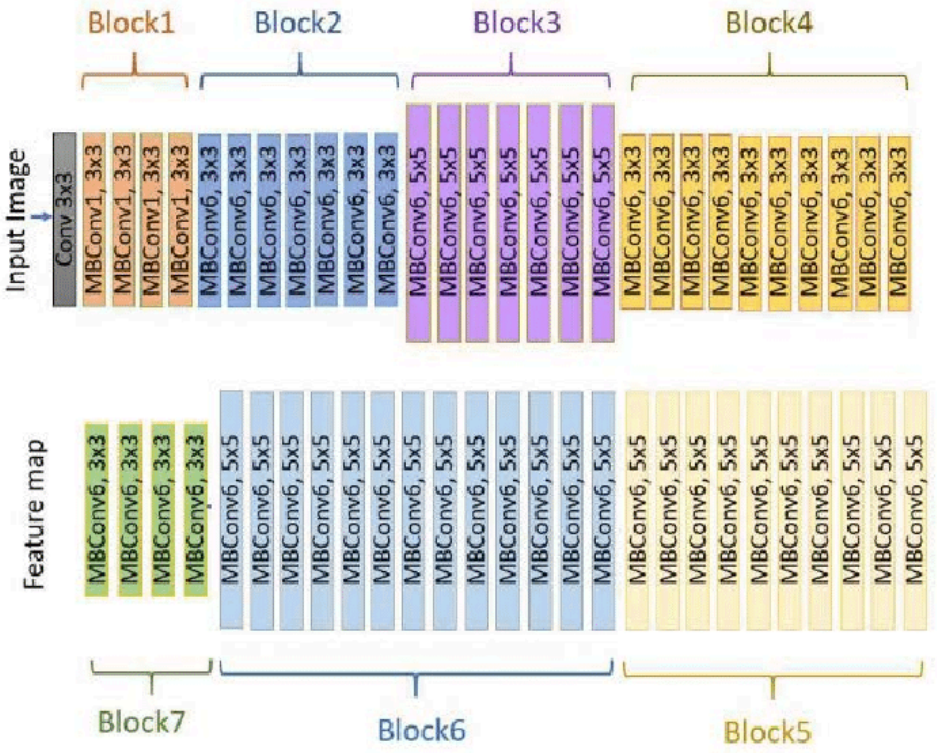


figure 11 : EfficientNet-b7 architecture

3. Vision Transformer Large (ViT-base, 384x384 & 224x224)

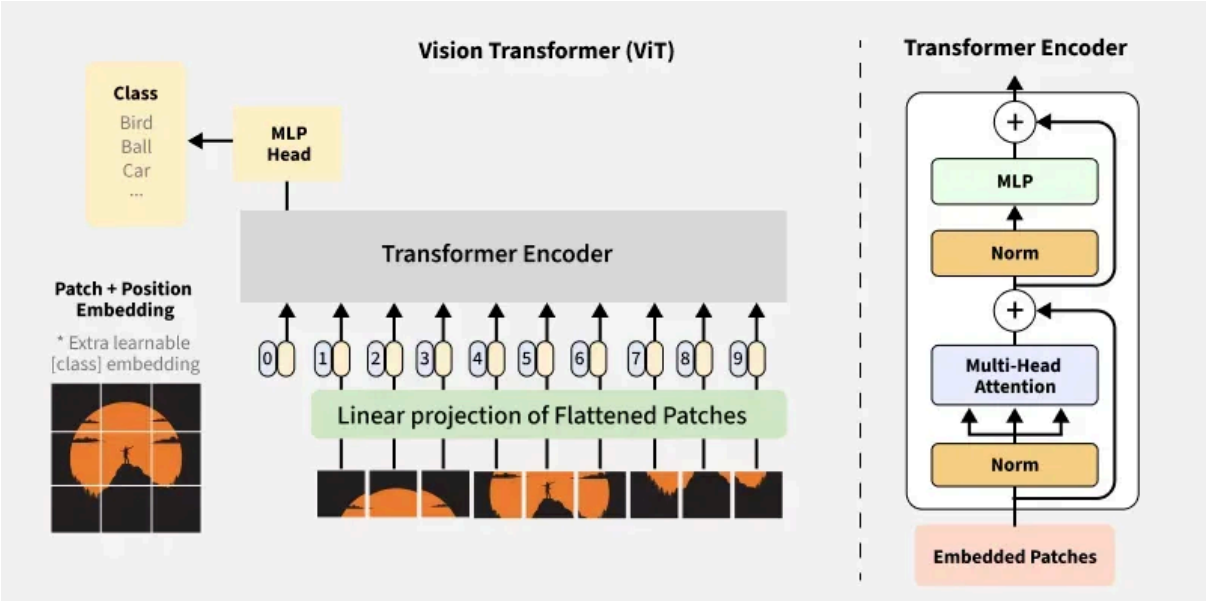


figure 10 : Vision Transformer architecture

4. Swin Transformer Large
5. Swin Transformer V2 Large

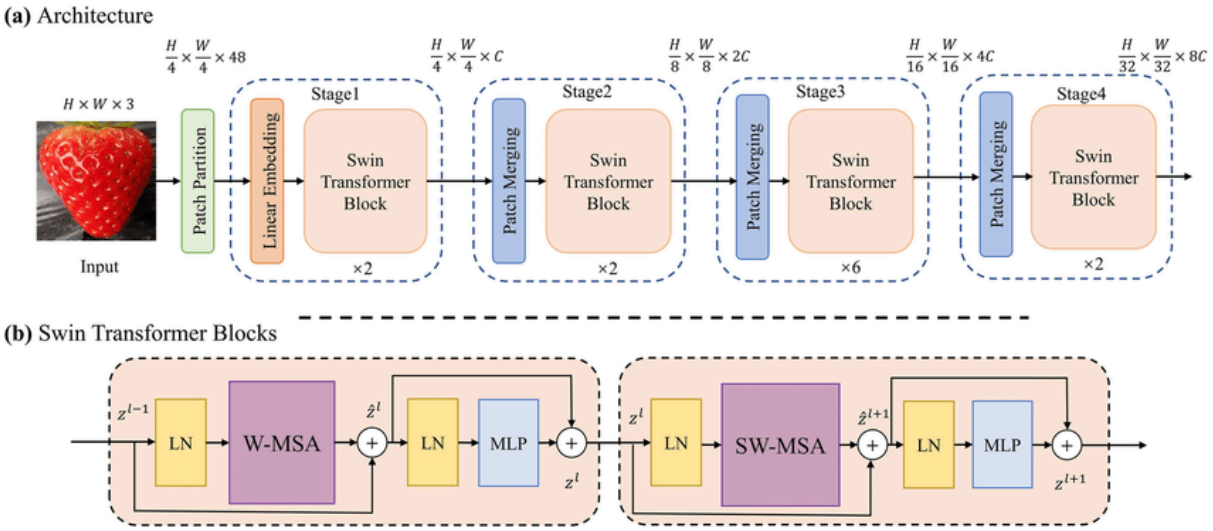


figure 11 : Swin Vision Transformer architecture

6. ViT-BiT CNN Hybrid
7. Custom Convolutional Network designed by Callan

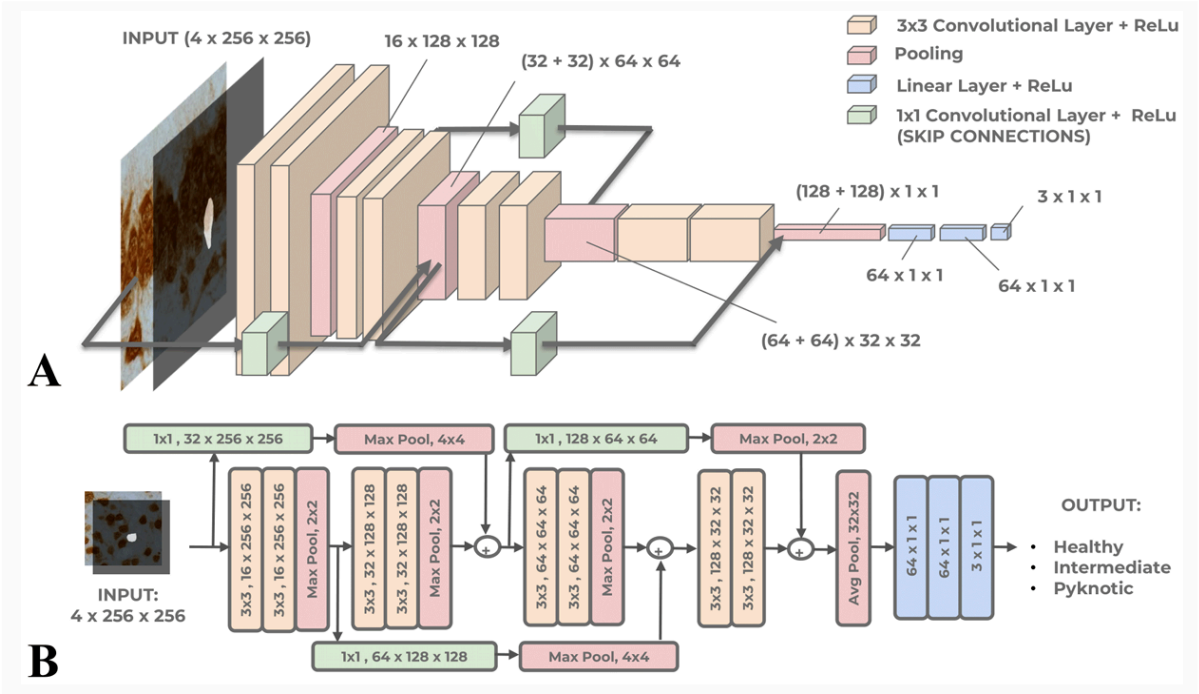


figure 12 : Callan's custom CNN architecture

c. Data augmentation

A difficult challenge in medical image analysis is severe class imbalance. Pyknotic cells are naturally less frequent in the provided data (cf. Dataset construction), leading to a dataset skewed heavily towards Normal and Intermediate cells. A model trained on such data would be biased and perform poorly on the minority class.
To guarantee that each fold of the 5-fold cross-validation was perfectly balanced during training and evaluation, I implemented a balancing algorithm. The process for each fold was as follows:

- The Scikit-learn StratifiedKFold function split the data into training and validation sets, initially preserving the original class ratios.

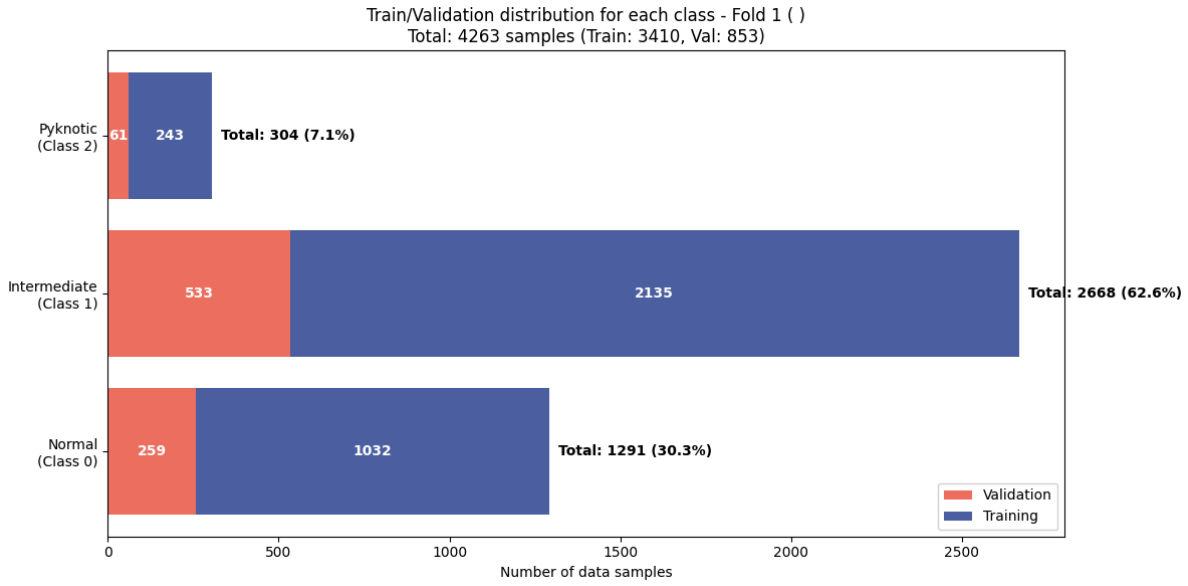


figure 13 : Data distribution before balancing

- A random subset of Normal and Intermediate cells was selected and removed from the validation set, exactly matching the count of the pyknotic cells. These removed samples were added to the training sets for these two classes.
- The cells in the two smallest training sets were then duplicated and underwent data augmentation until their count matched the number of cells in the most populated training set. This ensured the model was presented with an equal number of examples from each class during every epoch of training.

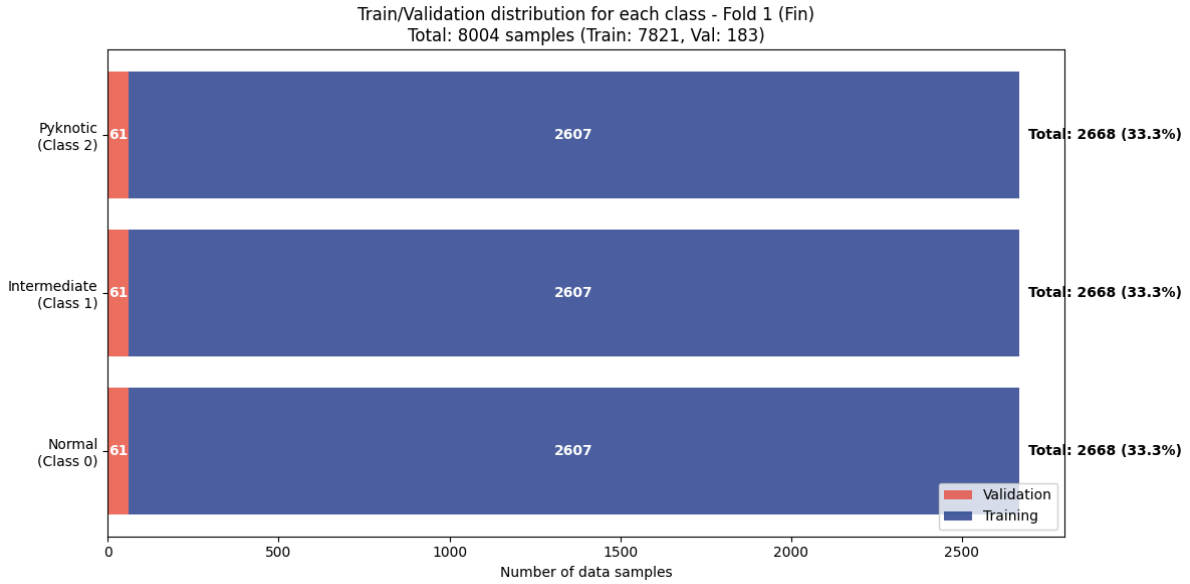


figure 14 : Data distribution after balancing

Duplicates detection was implemented using the hash functions of each augmented image, in order to verify if any of the augmented training and validation sets for each fold contained duplicates.

The augmentation occured during training with synchronized transformations to both images and masks. This augmentation, implemented using the ImgAug library, consisted of a suite of geometric and photometric transformations after normalization :

The geometric transformations include:

- Horizontal flips (50% probability) and vertical flips (20% probability).
- Affine Transformations applied with 50% probability, including scaling (0.8-1.2x), translation (±10%), rotation (±15°), shear (±8°), interpolation and pixel filling

The photometric transformations occurred randomly with 80% probability and a random selection from:

- Gaussian blur
- Additive gaussian noise
- Contrast adjustment
- Sharpening
- Brightness multiplication
- Channel Shuffling between the R, G and B channels

The final step of the data augmentation implements model-specific preprocessing of the image sizes that adapts the augmented data to the requirements of the different models, as different models require different input resolutions. (224×224 for some ViTs, 256×256 for CNNs, 384×384 for large transformers, 600×600 for EfficientNet)

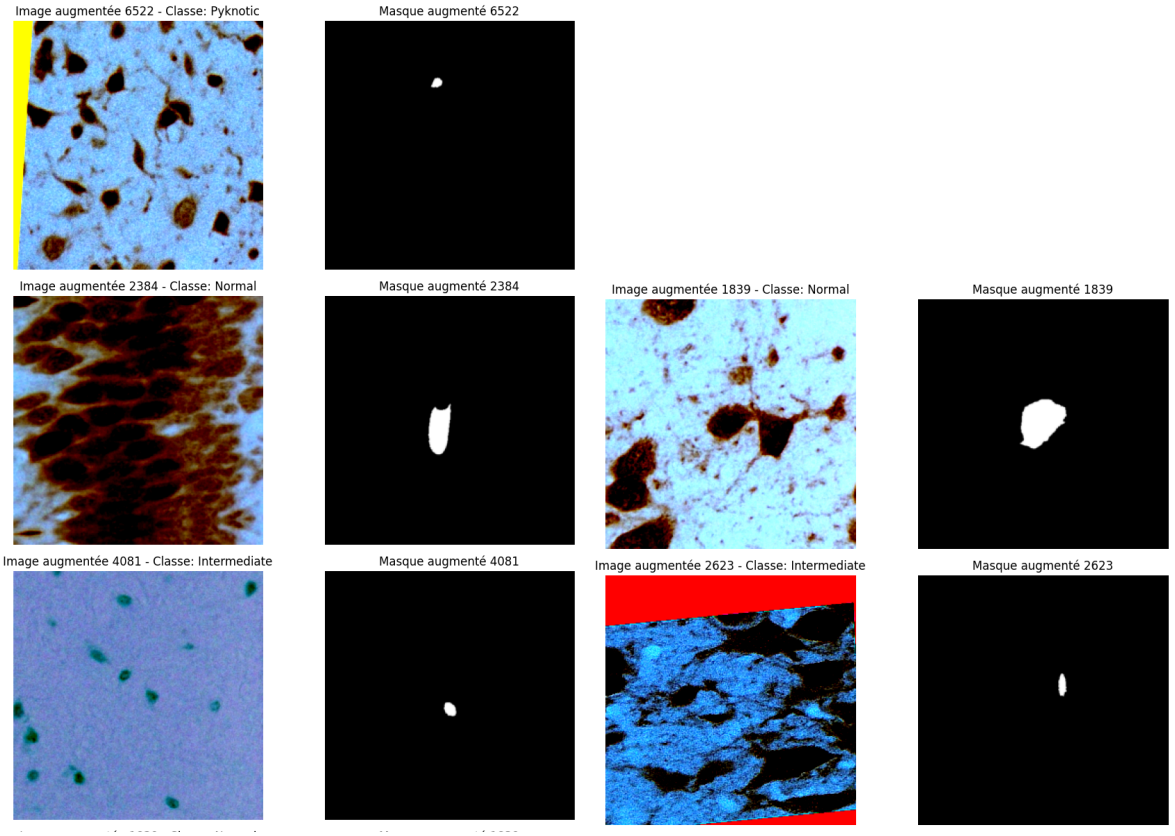


figure 15 : Examples of augmented 4-channel data, RGB channels to the right and masks to the left

d. Architecture modification of pre-trained models

A necessary modification of the first layers of all models was due to the fact that the segmentation and classification pipeline had been proven by Callan to work best with 4-channel inputs rather than 3-channel inputs. The 4-channel inputs are created by concatenating the 3-channel RGB image with its corresponding 1-channel segmentation mask.

All publicly available pre-trained models are exclusively designed to process 3-channel (RGB) input. Adapting this suite of state-of-the-art architectures to accept 4-channel data without discarding the valuable mask information required modifications to the initial input layers. Models pre-trained on the ImageNet dataset also have 1000 output classes.

An important challenge of this modification was to preserve the maximum amount of pre-trained weights of the original models while changing the number of input channels and output classes.

For the Hybrid ViT, a custom convolution layer with a kernel size of 1 and 4 input channels was added as the very first layer of the architecture. The kernel size of 1x1 was used so spatial revolution was not changed, and the weights of the initial 3 channels were copied onto

the weights of the custom convolution layer for the RGB channels. The weights of the fourth channel were initialized with a constant weight. The output classifier head was then replaced with a new classifier head (linear feed-forward layer) initialized with the right number of output classes.

For the Standard ViT (384x384 and 224x224) and PandaViT, the model's configuration data was directly modified to change the number of input channels to 4 and the number of output classes to 3 when retrieving the model.

For Swin and Swin-v2 Large, the model's configuration was adapted to a 4-channel input and 3-channel output directly when retrieving the model and a new convolutional layer with 4 input channels was created to replace the original patch embedding projection. The weights of the RGB channels were copied, and the fourth channel's weights were initialized with constants by copying the weights of the R channel.

For EfficientNet-b7 and ResNet-50, the first convolutional layer was replaced with a new convolutional layer initialized to accept 4 channels. The classifier head was replaced with a fully connected linear layer with 3 output classes.

e. Measures of comparison and evaluation of the models

The following measures of comparison and evaluation of the models were selected to provide an assessment of their performance, class-specific behavior, and computational efficiency.

The ROC (receiver operating characteristic) curve is a graphical representation of a classifier's performance across all possible decision thresholds. It plots the true positive rate against the false positive rate. The AUC (area under the receiver operating characteristic curve) quantifies the ability of the model to discriminate between classes, with an AUC of 1 indicating perfect discrimination, and an AUC of 0.5 being equivalent to a random guess. For each model, the AUC and the corresponding ROC curve were computed to evaluate the overall ability of the model to discriminate between classes, while the average ROC curve for each individual class was also computed to evaluate performance according to the different classes.

Accuracy, precision, recall, and F1-score were calculated for each class and then averaged using three averaging strategies: macro-average (computes the metric independently for each class and then averages, treating all classes equally), micro-average (adds together contributions of all classes to compute a global metric) and weighted-average (computes the metric for each class and averages them weighted by the number of true instances per class). Confusion matrices were generated to visualize the distribution of correct and incorrect predictions for each class.

The total number of parameters was recorded to quantify model complexity, and total training time as well as average inference time per sample were measured. To evaluate performance across folds, the average validation accuracy across all folds was calculated for each model. A paired t-test was performed to determine the statistical significance of differences in average validation accuracy between every pair of models.

f. Hyper-parameter choice

The performance of deep learning models is highly sensitive to the configuration of their training parameters, known as hyperparameters.

For each model, a review of published fine tuned models on Hugging Face was conducted. The focus was on identifying models that had been fine-tuned and had achieved high performance. A spreadsheet was created to compile the hyperparameters of these top-performing fine-tuned models. Key parameters recorded for each model included:

- Learning Rate
- Batch Size
- Number of Training Epochs
- Optimizer Type
- Drop out rate
- Weight Decay
- Learning Rate Scheduler Type and Warm-up Ratio
- Gradient Accumulation Steps

The primary criterion for selecting given hyperparameters was a reported validation accuracy exceeding 96% on its fine-tuned dataset. In cases where there was no model that met this threshold for a given architecture, the best available configuration was selected, even if its accuracy was slightly lower.

The hyper-parameters table can be found in the annex.

g. Training

The training phase of this study was conducted on New Zealand eScience Infrastructure (NeSI), New Zealand's national high-performance computing (HPC) facility. Training jobs were submitted through NeSI's Slurm workload manager. The GPUs used for training were 4 NVIDIA L4 units with around 23GB memory per GPU. K-fold cross-validation was used for training, as this study dealt with a limited dataset. It is a method for evaluating a model’s performance by splitting the dataset into K equally sized subsets, also called folds. The model is then trained K times, using one fold for validation and the remaining folds for training. By

sharing the data into 5 subsets of training and validation data and training the model 5 times on a different fold, the amount of available data for training and validation was more important than if a single train-validation split had been used. K-fold cross validation also provides a better estimation of performance and reduces variance due to random train-test splits.

B. Results

As mentioned above, the results of training so far comprise an array of different measures : The average validation accuracy across all folds, the inference time per image for each model after training (in milliseconds), the ROC curves for each model with their AUC (Area Under Curve), the ROC curves for each class for each model, confusion matrices for each model, and the accuracy, recall, F1 score and precision derived from the classification matrices, the total training time for each model and the statistical significance of the validation accuracy results per fold using a paired t-test.

The average validation accuracy results which can be found below give important insights :

1. The three best performing models are Swin-v2, the hybrid ViT and the custom CNN designed by Callan.
2. The standard ViT architectures are among the least performant models
3. The three top performing models are all of three completely different architectures (Swin vision transformer, ViT - CNN hybrid, CNN)
4. The validation accuracy difference between the top performing model and the worst performing model is 0.0526 (5,26%) which indicates that the most important performance difference between all models remains quite reasonable.

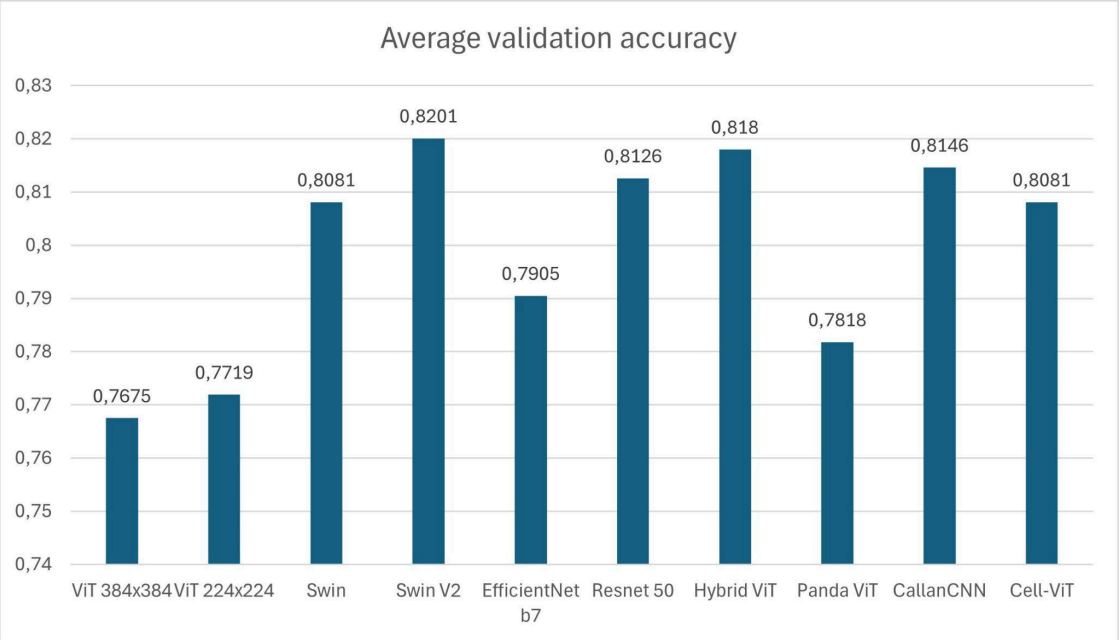


figure 16 : Average validation accuracy across all folds

Inference time shows important differences between models, showing that the Custom CNN model has a more promising performance to accuracy ratio than all other models, followed by the Hybrid ViT model :

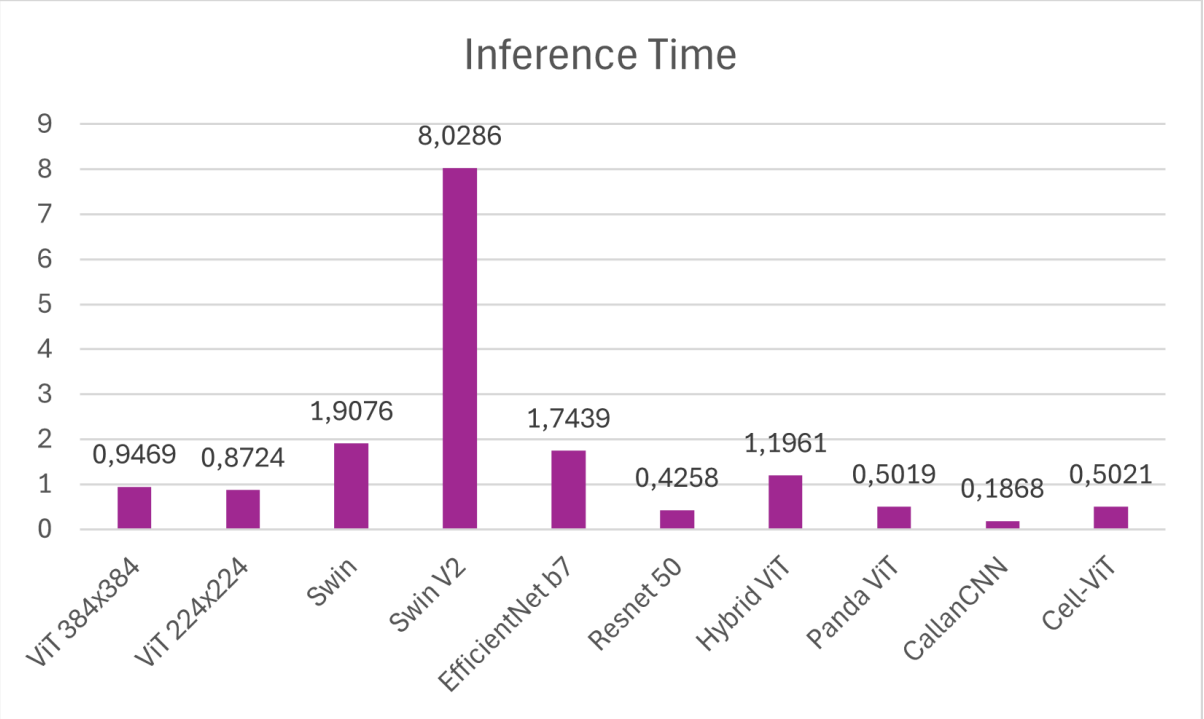


figure 17 : Inference time (milliseconds)

The confusion matrices for all models can be found below, showing the classification performance per class, according to the True Positives, False Positives, True Negatives and False Negatives. They show that very few models have shown incorrect classification between the Normal and the Pyknotic class, and that ResNet, EfficientNet, Swin and the Hybrid models show the best ability to correctly classify intermediate cells.

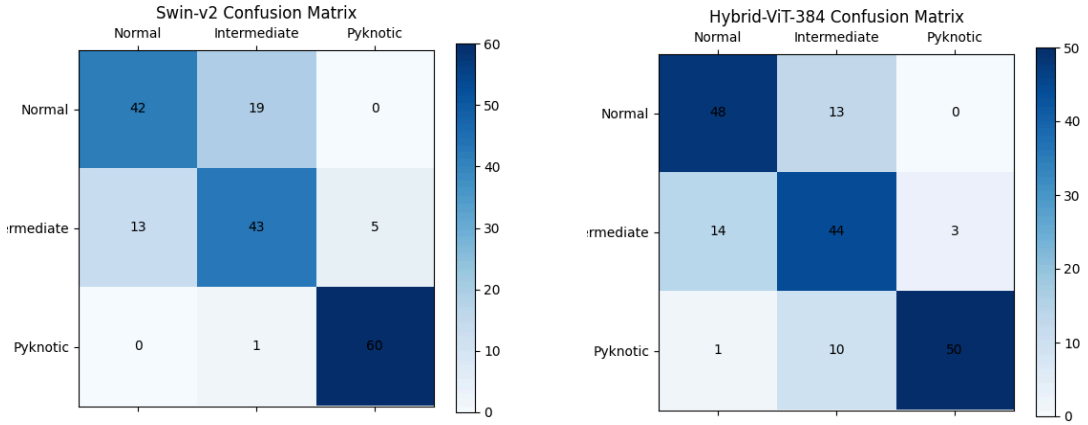


figure 18 : Confusion matrices of Swin-v2 and Hybrid-ViT models

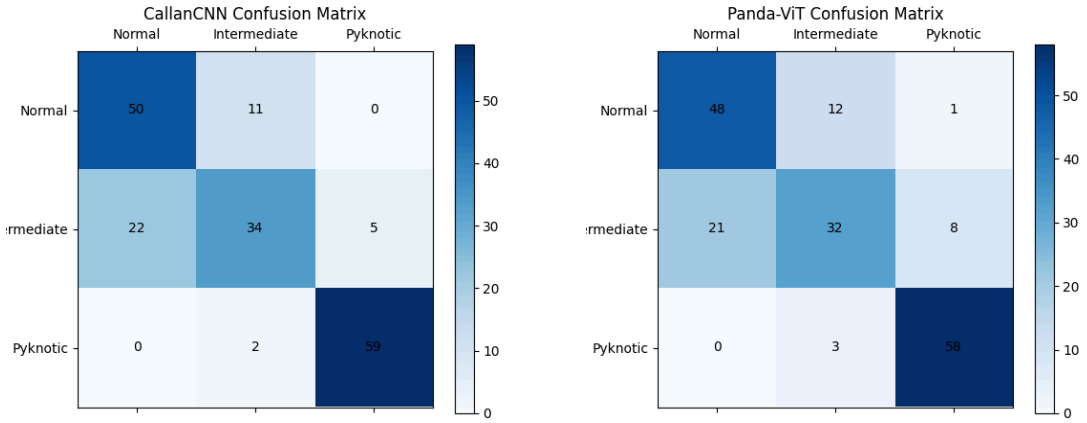


figure 19 : Confusion matrices of Callan's custom CNN and Panda-ViT models

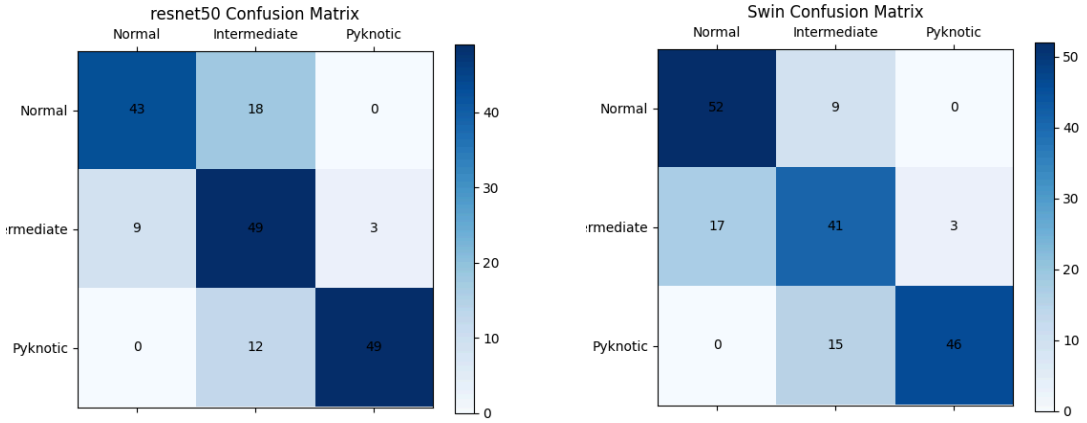


figure 20 : Confusion matrices of ResNet-50 and Swin models

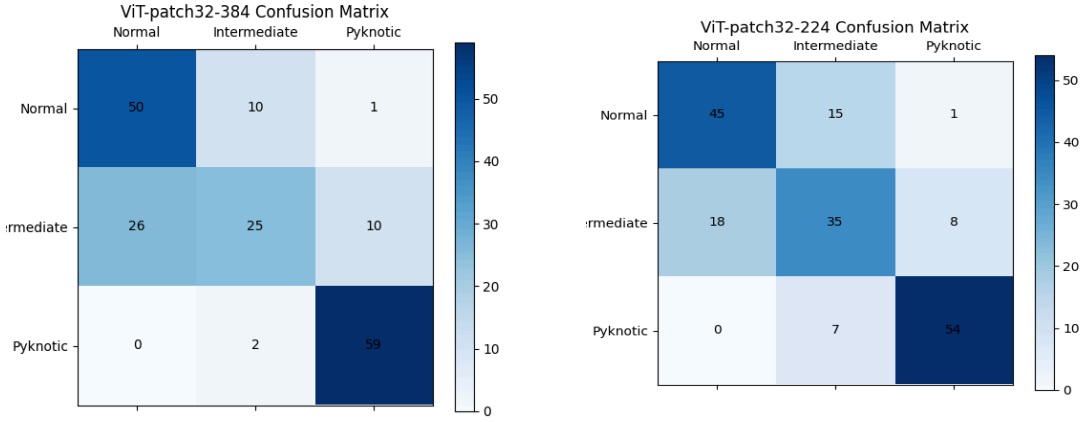


figure 21 : Confusion matrices of ViT (224x224) and (384x384) models

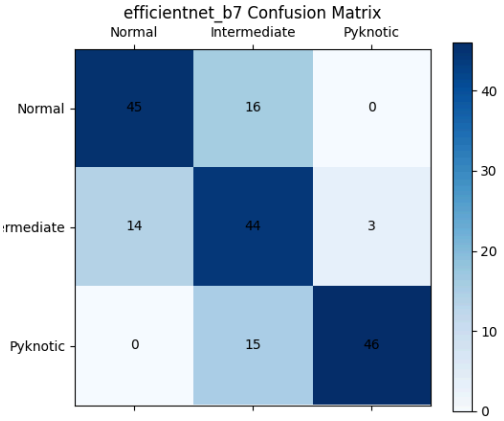


figure 22 : Confusion matrices of EfficientNet-b7

Accuracy, precision, recall and F1 score have been deduced from the confusion matrices. These metrics are calculated using the True Positive, False Positive, True Negative and False Negatives count for each class displayed by the confusion matrices.

True Positives (TP) : The positive predictions of one class made by the model that are correctly predicted. For example, when looking at the True Positives of the "Normal" class, we aim to count the number of times the model correctly predicts "Normal" for a cell that has been labeled as "Normal" in the dataset.

False Positives (FP) : The positive predictions of one class that are incorrectly predicted. For example, when looking at the False Positives of the "Normal" class, we aim to count the number of times the model incorrectly predicts "Normal" for a cell that has been labeled as "Pyknotic" (or "Intermediate") in the dataset.

True Negatives (TN) : The negative predictions of one class that are correctly predicted. For example, when looking at the True Negatives of the "Normal" class, we aim to count the number of times the model correctly predicts "Pyknotic" (or "Intermediate") for a cell that has been labeled as "Pyknotic" (or "Intermediate" respectively) in the dataset.

False negatives (FN) : The negative predictions of one class that are incorrectly predicted. For example, when looking at the False Negatives of the "Normal" class, we aim to count the number of times the model incorrectly predicts "Pyknotic" (or "Intermediate") for a cell that has been labeled as "Normal" in the dataset.

Accuracy = $\frac{correct\ predicted\ classifications}{total\ labeled\ classifications} = \frac{TP + TN}{TP + TN + FP + FN}$: The proportion of all (positive and negative) classifications that were correctly predicted.

Precision = $\frac{correct\ positive\ predicted\ classifications}{total\ positive\ classifications} = \frac{TP}{TP + FP}$: The proportion of positive classifications for each class that were identified correctly.

Recall = $\frac{correct\ positive\ predicted\ classifications}{total\ positive\ labels} = \frac{TP}{TP+FN}$: The proportion of all positive data that were correctly classified positive.

F1-score $= 2 * \frac{precision * recall}{precision + recall} = \frac{2TP}{2TP+FP+FN}$: The harmonic mean of precision and recall.

Accuracy, precision, recall and F1-score for each class can be found below:

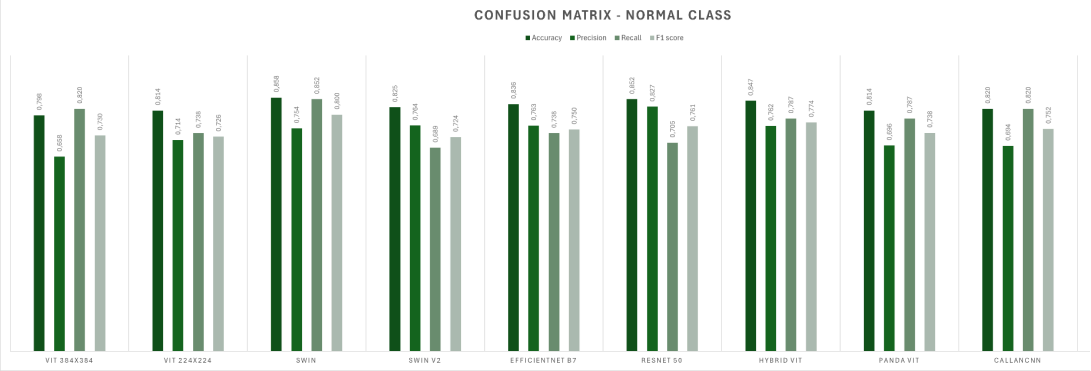


figure 23 : F1-score, accuracy, precision and recall for the normal class

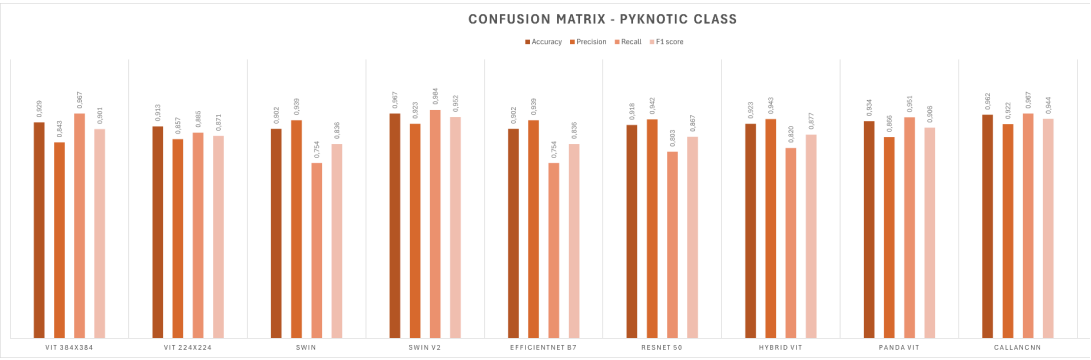


figure 24 : F1-score, accuracy, precision and recall for the normal class

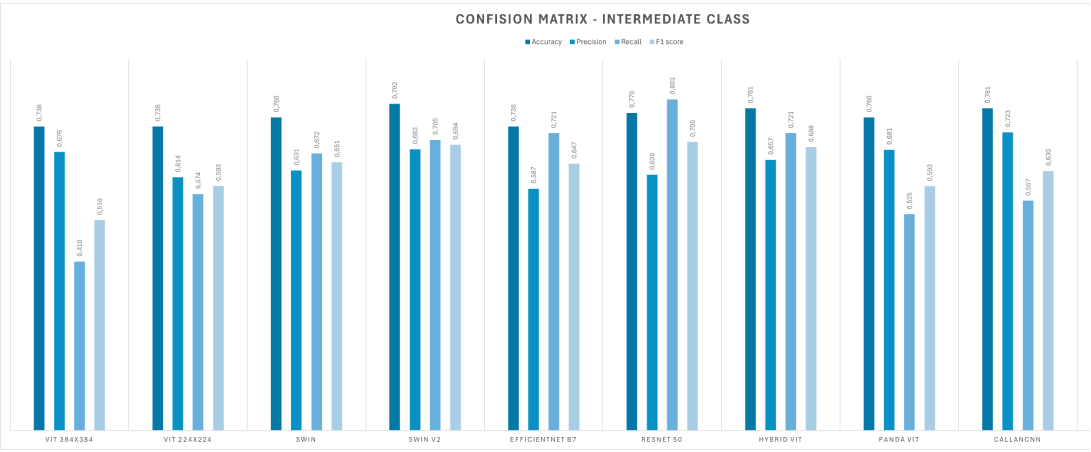


figure 25 : F1-score, accuracy, precision and recall for the normal class

The micro, macro and weighted average of these metrics for all classes can be found below.

figure 26 : F1-score, accuracy, precision and recall macro-averaged for all classes

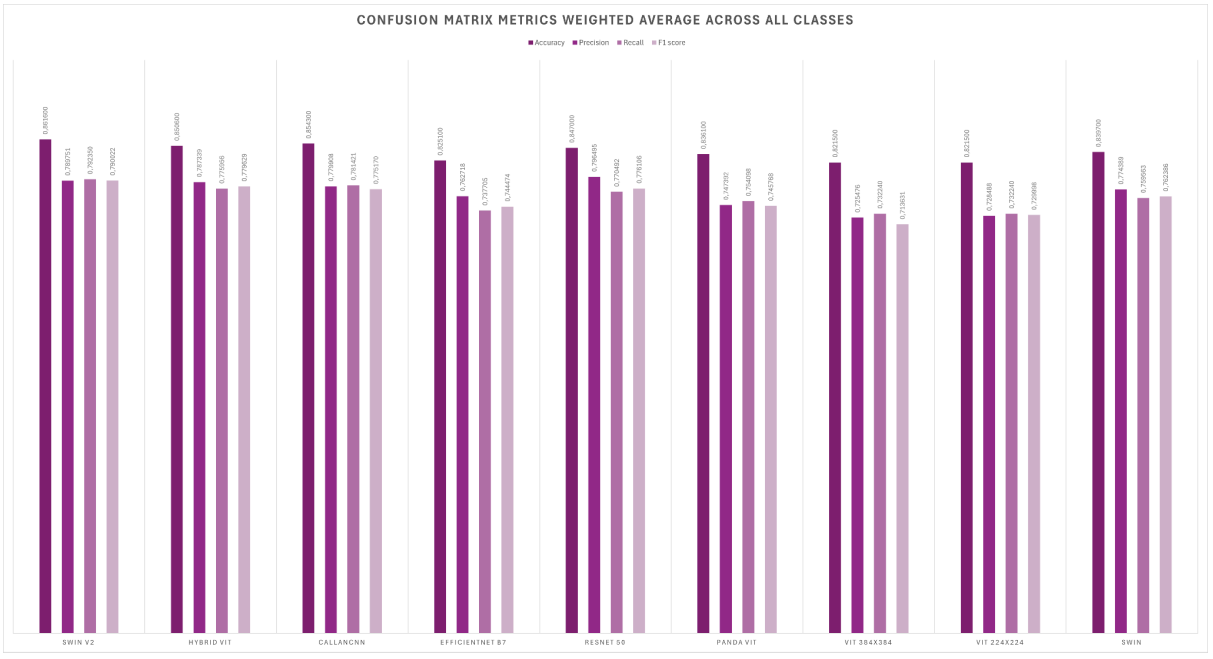


figure 27 : F1-score, accuracy, precision and recall weighted-averaged for all classes

figure 28 : F1-score, accuracy, precision and recall micro averaged for all classes

They show that across all averaging methods, Swin-V2, the custom CNN and the Hybrid ViT consistently achieve the highest scores.

ROC Curves were also obtained. An average ROC with AUC was obtained for each model, and class specific ROCs were computed for the validation data across folds and averaged in order to yield an average ROC per class per model.

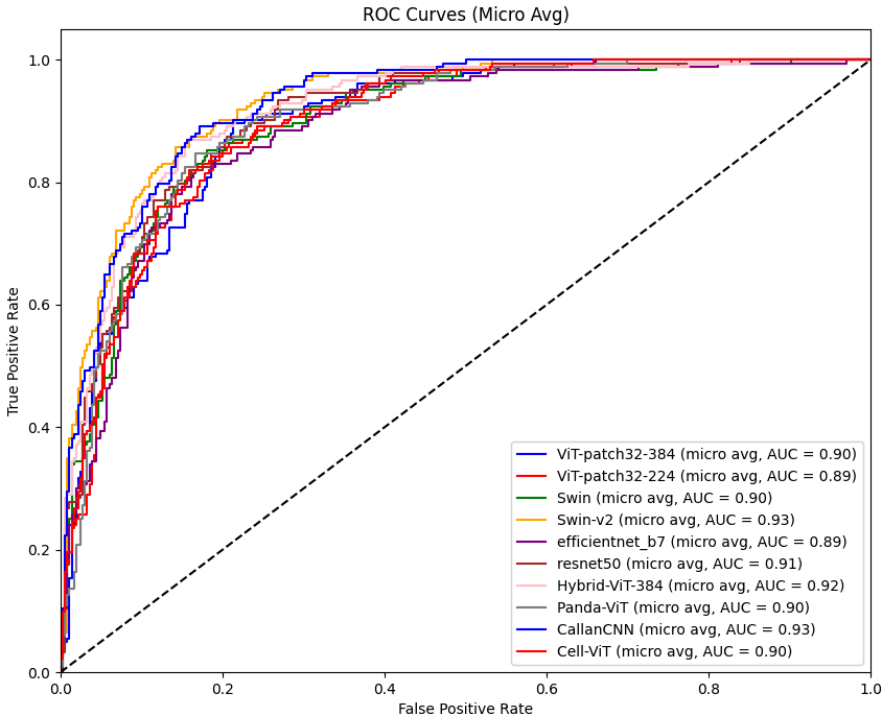


figure 29 : Average ROC Curves for each model and AUC

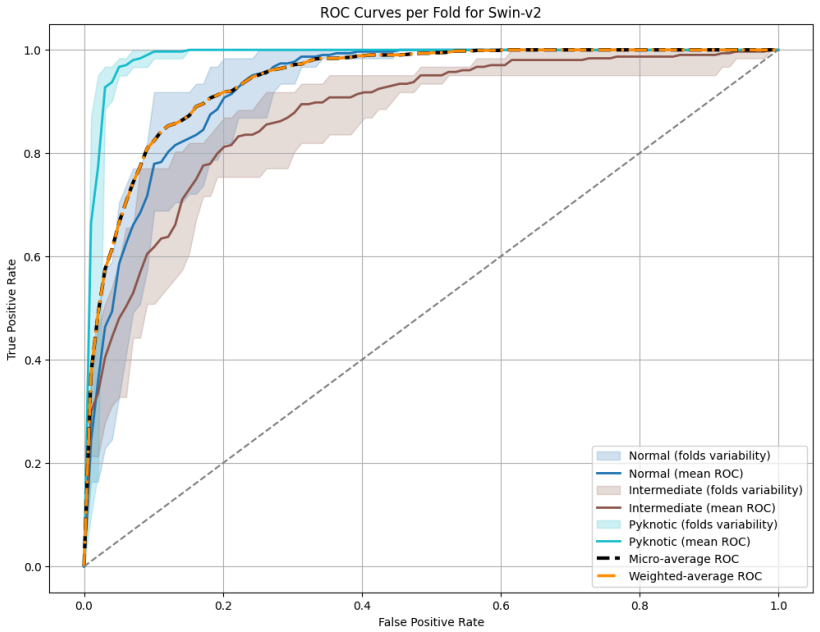


figure 30 : Average ROC Curve, and ROC curves per class and fold for Swin-v2 model

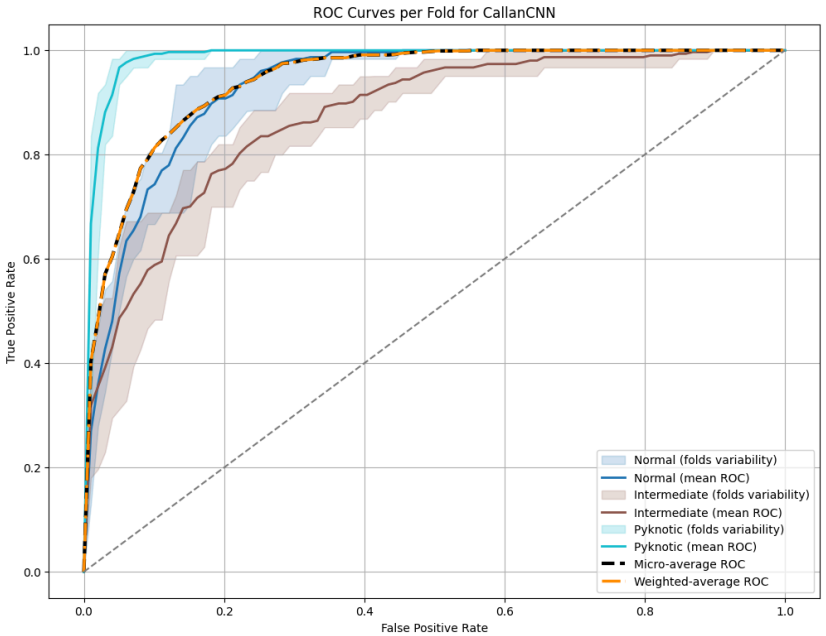


figure 31 : Average ROC Curve, and ROC curves per class and fold for Callan's custom CNN model

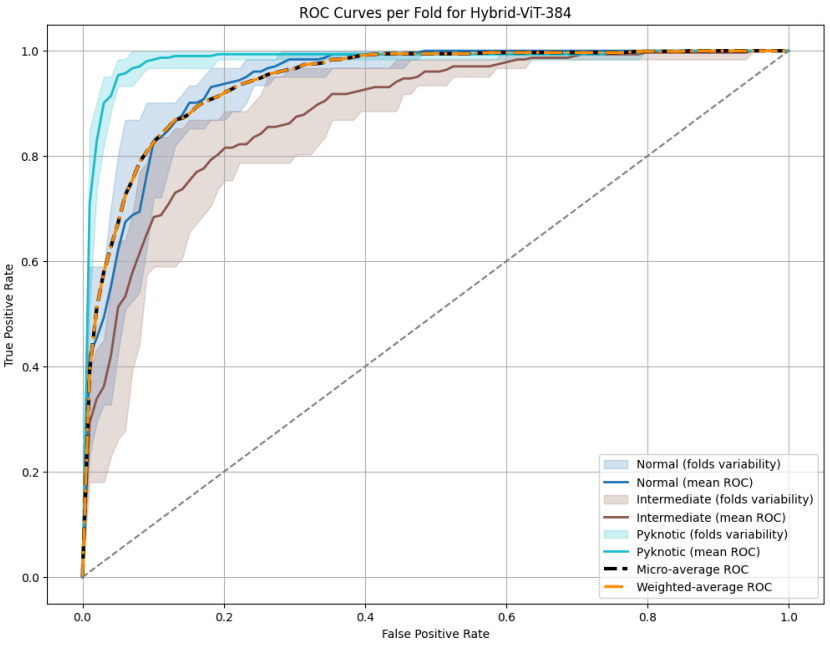


figure 32 : Average ROC Curve, and ROC curves per class and fold for the Hybrid-ViT model

They show that on average, the Pyknotic class is best classified, followed by the normal class. The intermediate class consistently shows a lower true positive to false positive ratio, meaning that the models have more difficulty distinguishing between actual positives and incorrectly identified positive cases compared to the two other classes. This is due to the fact that the intermediate class has characteristics that overlap with the other classes, making it harder for the models to classify accurately. This is supported by the F1-score, accuracy, precision and recall in figures 23, 24 and 25 above.

The AUC scores vary from 0.89 to 0.93, suggesting that all models demonstrate a consistent and comparable performance.

A paired t-test was conducted to determine whether differences in model performance were statistically significant. Validation accuracies from each fold were used as the performance metric. The significance level was set to $\alpha = 0.05$. The t-statistic and p-values are reported below:

Model	ViT-patch384	ViT-patch224	Swin	Swin-v2	EfficientNetB7	ResNet50	Hybrid-ViT	Panda-ViT	CallanCNN
ViT-patch384	0.000	-1.633	-5.724	-4.681	-1.596	-3.794	-4.872	-2.219	-8.917
ViT-patch224	1.633	0.000	-4.401	-3.755	-1.128	-3.197	-4.171	-1.503	-5.800
Swin	5.724	4.401	0.000	-0.818	1.654	-0.418	-0.923	3.461	-0.692
Swin-v2	4.681	3.755	0.818	0.000	2.115	0.655	0.222	2.430	0.845
EfficientNetB7	1.596	1.128	-1.654	-2.115	0.000	-1.644	-1.989	0.505	-2.002
ResNet50	3.794	3.197	0.418	-0.655	1.644	0.000	-1.826	2.548	-0.185
Hybrid-ViT	4.872	4.171	0.923	-0.222	1.989	1.826	0.000	3.069	0.363
Panda-ViT	2.219	1.503	-3.461	-2.430	-0.505	-2.548	-3.069	0.000	-3.038
CallanCNN	8.917	5.800	0.692	-0.845	2.002	0.185	-0.363	3.038	0.000

figure 33 : T values for each model pair

Model	ViT-patch384	ViT-patch224	Swin	Swin-v2	EfficientNetB7	ResNet50	Hybrid-ViT	Panda-ViT	CallanCNN
ViT-patch384	0.000	0.178	0.005	0.009	0.186	0.019	0.008	0.091	0.001
ViT-patch224	0.178	0.000	0.012	0.020	0.322	0.033	0.014	0.207	0.004
Swin	0.005	0.012	0.000	0.459	0.173	0.697	0.408	0.026	0.527
Swin-v2	0.009	0.020	0.459	0.000	0.102	0.548	0.835	0.072	0.446
EfficientNetB7	0.186	0.322	0.173	0.102	0.000	0.176	0.118	0.640	0.116
ResNet50	0.019	0.033	0.697	0.548	0.176	0.000	0.142	0.063	0.862
Hybrid-ViT	0.008	0.014	0.408	0.835	0.118	0.142	0.000	0.037	0.735
Panda-ViT	0.091	0.207	0.026	0.072	0.640	0.063	0.037	0.000	0.038
CallanCNN	0.001	0.004	0.527	0.446	0.116	0.862	0.735	0.038	0.000

figure 34 : P values for each model pair

Other metrics can be found in the annex, such as the total training time per model and all the ROC Curves.

For subsequent training, the code will be modified to include additionally outputting the following information, which lacked in the first training :

- The evolution of the validation accuracy per epoch during each fold
- The evolution of the training loss and validation loss during each fold
- A calculation of the mean and standard deviation of validation accuracies
- The inference time every X epochs, including the inference time per batch using the full classification-segmentation pipeline for each model
- Memory consumption
- An representation of the validation accuracy for the six different brain regions

C. Discussion

As said above, the best performing models are the Swin-v2, the Hybrid-CNN, and the custom CNN models, with 82.01%, 81.80% and 81.46% validation accuracy respectively. These models have very different architectures, and their very similar performances indicate that the performance limitations lie rather within the dataset than the model architectures.

Based on the statistical comparison tables above (t values and p values), Swin-v2 and the custom CNN emerge as the strongest performers. Swin-v2 demonstrates a consistent positive performance differential against nearly all other models, however its advantages over all other models outside of the standard ViTs are not statistically significant. This also applies to the custom CNN model, whose advantages are not statistically significant compared to other models outside of the standard ViTs and the Panda-ViT.

VIT-patch384 and Panda-VIT are consistently outperformed by the majority of other models. VIT-patch384, in particular, shows statistically significant losses against almost every other model in the comparison. Panda-VIT is also consistently beaten, with its performance compared to top models like Swin and Callan's custom CNN being statistically significant.

The underperformance of ViT models can be explained by the size limitation of the fine tuning dataset, as transformer models require a much higher amount of data in order to infer robust features and global relationships. The t and p values tables show that this underperformance is indeed statistically significant when comparing their average validation accuracies with all other models, except EfficientNet-b7 and the Panda-ViT.

Standard ViTs split the image into patches and linearly embed them. These patch embeddings, plus positional information, are fed directly into a standard Transformer encoder that uses global self-attention from the very first layer. This feature extraction favorises global features, which means that standard ViT models will not naturally favor local rather than global relationships. This could explain the performance difference as it means that a ViT model requires a larger amount of data in order to perform better on a vision task.

On the contrary, the Swin, CNN and hybrid models' architectures favor local relationships in feature extraction. In the Swin model, the self-attention mechanism also known as shifted window attention is limited to n neighboring patches for each patch of an image, and CNN models rely on kernels of fixed sizes to extract feature maps. The incorporation of local spatial biases into a transformer architecture seems to be the reason the Swin-v2 and the Hybrid ViT-CNN architecture perform best.

According to the statistical metrics, the performance gap between VIT-patch224 and VIT-patch384 is both consistent and significant, suggesting that for this specific task, the smaller patch size and its implications for input resolution are more effective. Furthermore, the newer Swin-v2 model shows a consistent, if not always significant, improvement over its predecessor, Swin.
The confusion matrices indicate that the Swin-v2 transformer performs extremely well at identifying pyknotic cells and normal cells. In fact, it is the best model for identification of pyknotic cells, closely followed by the custom CNN and the standard 384x384 ViT. It shows no confusion between the two classes, as well as a higher than average ability to identify intermediate cells correctly (41 correctly identified intermediate cells, 26 cells confused between normal and intermediate, as well as 18 cells confused between intermediate and pyknotic).
The custom CNN, Swin and standard 384x384 ViT are the best performers in normal cell identification, with the Swin model being most efficient at identifying intermediate cells out of the three models. The standard 384x384 ViT's global performance is highly undermined by its restricted ability to identify intermediate cells correctly, with 25 correctly identified for 36 incorrectly identified intermediate cells.
The best models for intermediate cells identification are ResNet-50 (49 correctly identified, 12 incorrectly identified), Swin-v2 (43 correctly identified, 18 incorrectly identified) EfficientNet-b7 (44 correctly identified, 17 incorrectly identified) and the Hybrid-ViT (44 correctly identified). We can thus conclude that transformer architectures are not of any advantage when facing the challenge of more accurately classifying intermediate cells. These results show that CNN architectures perform best in that regard, on par with hybrid transformer architectures.

The differences in inference time of the Swin (8.0286 ms), hybrid (1.1961 ms) and custom CNN (0.1868) models are much more significant than the differences in average validation accuracies, making the custom CNN model the model with the best accuracy-to-speed ratio, and the most optimal choice for deployment in a real-world application setting.

As said above, the AUC scores vary from 0.89 to 0.93. This suggests that all models demonstrate a consistent and comparable performance. The highest variation being of only 0.04 suggests that no single model considerably outperforms the others.
This offers an interesting insight : given the fact that different model architectures have been trained and fine-tuned on the same datasets (standard ViTs, hybrid ViT and Swin-v2 trained on ImageNet-21k, Swin, EfficientNet-b7, ResNet-50 trained on ImageNet-1k), we can conclude that the model architecture only has a limited impact on its performance on this given fine-tuning dataset. Despite their architectural diversity, the convergence of performance metrics suggests that the primary bottleneck lies in dataset characteristics rather than model capacity or architectural choices.
This may be due to the quantity of data available, sub-par annotation quality, and the important class imbalance.

This is supported by the classification metrics inferred from the confusion matrices (figure 18-21). Across all models, accuracy is consistently higher than precision, recall and F1-score in the weighted average. This suggests class imbalance and artificial inflation of accuracy. The differences in results between the averaging methods also suggest class imbalance, as the macro- and micro-averaging results quite significantly differ.
The Swin-v2, ResNet-50 and custom CNN architectures demonstrate the most consistency across averaging methods, which goes to show that they handle class imbalance best. Contrarily, EfficientNet-b7, and the standard ViTs show much more varying performances across different averaging methods.

D. Conclusion

In conclusion, the Swin-v2, Hybrid-ViT-CNN, and custom CNN models emerge as the best-performing architectures, achieving validation accuracies of 82.01%, 81.80%, and 81.46% respectively. Statistical comparisons confirm this claim. While Swin-v2 consistently outperforms most other models, its advantage over non-standard ViTs is not statistically significant, a pattern also observed for the custom CNN. Standard ViTs, particularly VIT-patch384, are outperformed by most other models, likely due to the limited size of the fine-tuning dataset, which is insufficient for transformers to fully exploit global feature relationships. Architectures favoring local feature extraction such as Swin, CNNs, and hybrid models demonstrate superior performance, highlighting the importance of local spatial biases for this task.

Analysis of confusion matrices confirms these trends too. Swin-v2 excels at identifying pyknotic and normal cells, while ResNet-50, Swin-v2, EfficientNet-b7, and the Hybrid-ViT perform best for intermediate cells, suggesting that transformer architectures do not offer inherent advantages for more challenging cell classes. In terms of inference speed, the custom CNN achieves the highest accuracy-to-speed ratio, making it the most practical choice for deployment.

VII. Critical analysis

A. Ethical aspects regarding the dataset

The histological images central to this project are the result of in vivo experiments on fetal sheep. While all experiments were ethically approved (see annex) by the Animal Ethics Committee of The University of Auckland under the New Zealand Animal Welfare Act, one might still question the ethical implications of artificially induced hypoxic-ischemia on sheep fetuses in the womb of their mothers.

B. Work methodology

Due to time constraints, some parts of the methodology were perhaps not delved into with enough detail, such as the hyper-parameter selection. It would perhaps have been more interesting to compute hyper-parameter searches before training.

The most discutable part of the work methodology however is the manual extension of the dataset. Callan's custom CNN had proven a 93% validation accuracy during his study on the initial 1500 cells dataset. This study has shown his model to have a 81.46% validation accuracy, due to the lack of balance in the extended dataset and surely due to some extent to incorrect labeling, as I performed the labeling myself without any prior formation.

C. Assessment

This internship has enabled me to further the technical skills that the HEALTH specialization had started teaching me during the school year, particularly concerning artificial intelligence and neural networks. It has also enabled me to continuously work in autonomy and self-manage my tasks and my own progress, which was a very important skill to acquire in order to deliver the work in a timely manner.

One of the challenges I faced was the unpredictability of the time it would take to carry out the very first training. Getting used to the NeSI interface and the fact that I did not know how long the training would take ended up costing me a few days worth of training time, as the training time was significantly longer than I had initially expected. Making any single change, no matter how small it was, to the training code once training was done meant losing considerable time.

VIII. Conclusions and perspectives

A. Journal paper

This work is currently being drafted into a research paper for the journal Bioengineering. The manuscript will provide all relevant information regarding the methodology and the findings presented above.

B. Professional project

This internship has been an experience which definitely solidified my desire to pursue a career in the field of artificial intelligence applied to healthcare. I immensely enjoyed learning more and manipulating deep learning models, and would like to continue in this field. I also enjoyed the research environment of the start-up, and am considering pursuing a PhD position in France.

IX. Bibliography

[1] Graham EM, Ruis KA, Hartman AL, Northington FJ, Fox HE. A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy. Am J Obstet Gynecol. 2008;199(6):587–595. doi:10.1016/j.ajog.2008.06.094

[2] Chalak LF, Kaiser JR, Sanchez PJ, et al. Prospective research in infants with mild encephalopathy identified in the first six hours of life: neurodevelopmental outcomes at 18–22 months. Early Hum Dev. 2011;87(9):705–710. doi:10.1016/j.earlhumdev.2011.05.010

[3] Vannucci RC, Perlman JM. Interventions for perinatal hypoxic-ischemic encephalopathy. Pediatrics. 1997;100(6):1004–1014.

[4] Long M, Brandon DH. Induced hypothermia for neonates with hypoxic-ischemic encephalopathy. J Obstet Gynecol Neonatal Nurs. 2007;36(3):293–298. doi:10.1111/j.1552-6909.2007.00153.x

[5] Shankaran S. Neonatal encephalopathy: treatment with hypothermia. J Neurotrauma. 2009;26(3):437–443. doi:10.1089/neu.2008.0678

[6] Davidson JO, Wassink G, Yuill CA, Zhang FG, Bennet L, Gunn AJ. How long is too long for cerebral cooling after ischemia in fetal sheep? J Cereb Blood Flow Metab. 2015;35(5):751–758. doi:10.1038/jcbfm.2014.247

[7] Jacobs SE, Berg M, Hunt R, Tarnow-Mordi WO, Inder TE, Davis PG. Cooling for newborns with hypoxic ischaemic encephalopathy. Cochrane Database Syst Rev. 2013;(1):CD003311. doi:10.1002/14651858.CD003311.pub3

[8] Davidson JO, Wassink G, van den Heuij LG, Bennet L, Gunn AJ. Therapeutic hypothermia for neonatal hypoxic–ischemic encephalopathy – where to from here? Front Neurol. 2015;6:198. doi:10.3389/fneur.2015.00198

[9] Edwards AD, Brocklehurst P, Gunn AJ, Halliday H, Juszczak E, Levene M, et al. Neurological outcomes at 18 months of age after moderate hypothermia for perinatal hypoxic ischaemic encephalopathy: synthesis and meta-analysis of trial data. BMJ. 2010;340:c363. doi:10.1136/bmj.c363

[10] Leung T, Weerapong P, Fraser M, et al. Neuroprotection by therapeutic hypothermia: mechanisms and clinical applications. Neurol Sci. 2020;41:313–323. doi:10.1007/s13760-020-01308-3

[11] Stoll BJ, Hansen NI, Bell EF, et al. Neonatal outcomes of extremely preterm infants from the NICHD Neonatal Research Network. Pediatrics. 2010;126(3):443–456. doi:10.1542/peds.2009-2959

[12] Loomes C, Davidson JO, Turatsinze Q, Gunn AJ, Bennet L, Abbasi H. Automated cell quantification in hypoxic-ischemic fetal sheep brain histology: a two-step segmentation and classification pipeline. IEEE Trans Biomed Eng. 2022;69(11):3394–3404. doi:10.1109/TBME.2022.3143456

[13] Volpe JJ. Encephalopathy of prematurity includes neuronal abnormalities. Pediatrics. 2005;116(1):221–225. doi:10.1542/peds.2005-0199

[14] Xiao T, Singh M, Mintun E, Darrell T, Dollár P, Girshick R. Early convolutions help transformers see better. Adv Neural Inf Process Syst. 2021;34:30392–30400.

[15] Dai Z, Liu H, Le QV, Tan M. CoAtNet: marrying convolution and attention for all data sizes. arXiv preprint. 2021. arXiv:2106.04803.

[16] Davies A, Wassink G, Bennet L, Gunn AJ, Davidson JO. Can we further optimize therapeutic hypothermia for hypoxic-ischemic encephalopathy? Neural Regen Res. 2019;14(10):1678–1683. doi:10.4103/1673-5374.257512

[17] Lakadia MJ, Abbasi H, Gunn AJ, Unsworth CP, Bennet L. Examining the effect of $MgSO_4$ on sharp wave transient activity in the hypoxic-ischemic fetal sheep model. Annu Int Conf IEEE Eng Med Biol Soc. 2016;2016:908-11. doi:10.1109/EMBC.2016.7590848. PMID: 28268471.

[18] Abbasi H, Bennet L, Gunn AJ, Unsworth CP. Identifying stereotypic evolving micro-scale seizures (SEMS) in the hypoxic-ischemic EEG of the pre-term fetal sheep with a wavelet type-II fuzzy classifier. Annu Int Conf IEEE Eng Med Biol Soc. 2016;2016:973-6. doi:10.1109/EMBC.2016.7590864. PMID: 28268486.

[19] Abbasi H, Bennet L, Gunn AJ, Unsworth CP. Latent phase detection of hypoxic-ischemic spike transients in the EEG of preterm fetal sheep using reverse biorthogonal wavelets and fuzzy classifier. Int J Neural Syst. 2019;29(10):1950013. doi:10.1142/S0129065719500138. PMID: 31184228.

[20] Abbasi H, Bennet L, Gunn AJ, Unsworth CP. 2D wavelet scalogram training of deep convolutional neural network for automatic identification of micro-scale sharp wave biomarkers in the hypoxic-ischemic EEG of preterm sheep. Annu Int Conf IEEE Eng Med Biol Soc. 2019;2019:1825-8. doi:10.1109/EMBC.2019.8857665. PMID: 31946252.

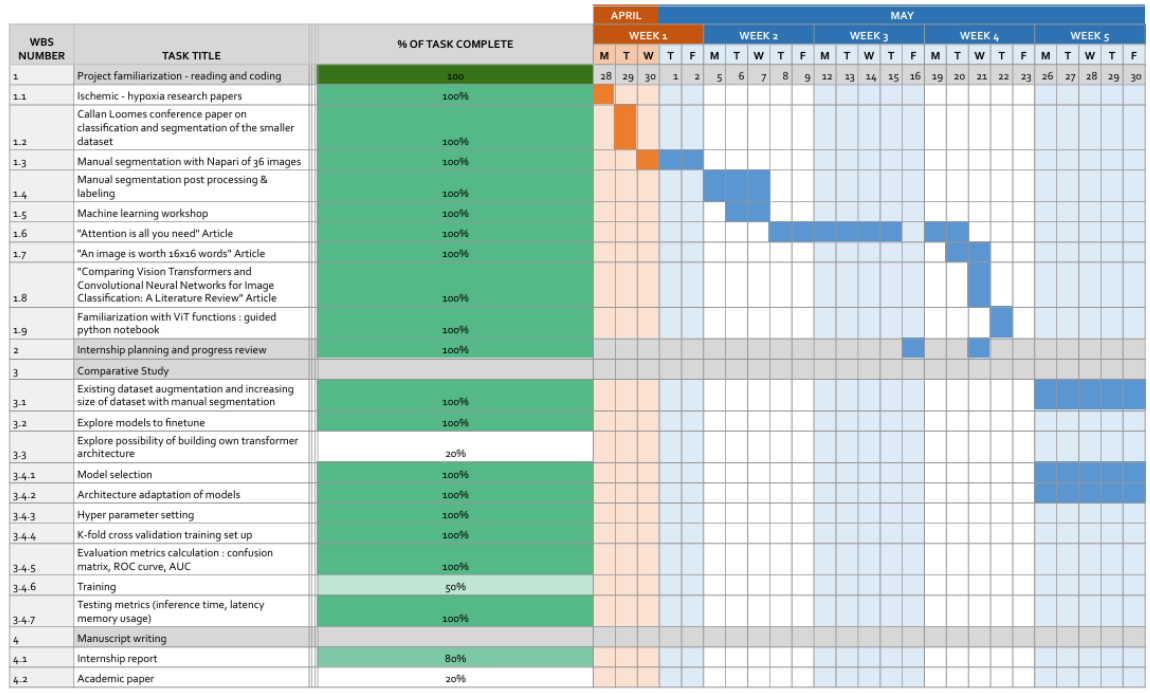
[21] Abbasi H, Davidson JO, Dhillon SK, Zhou KQ, Wassink G, Gunn AJ, Bennet L. Deep learning for generalized EEG seizure detection after hypoxia-ischemia: preclinical validation. Bioengineering (Basel). 2024;11(3):217. doi:10.3390/bioengineering11030217. PMID: 38534490; PMCID: PMC10968073.

X. Annex

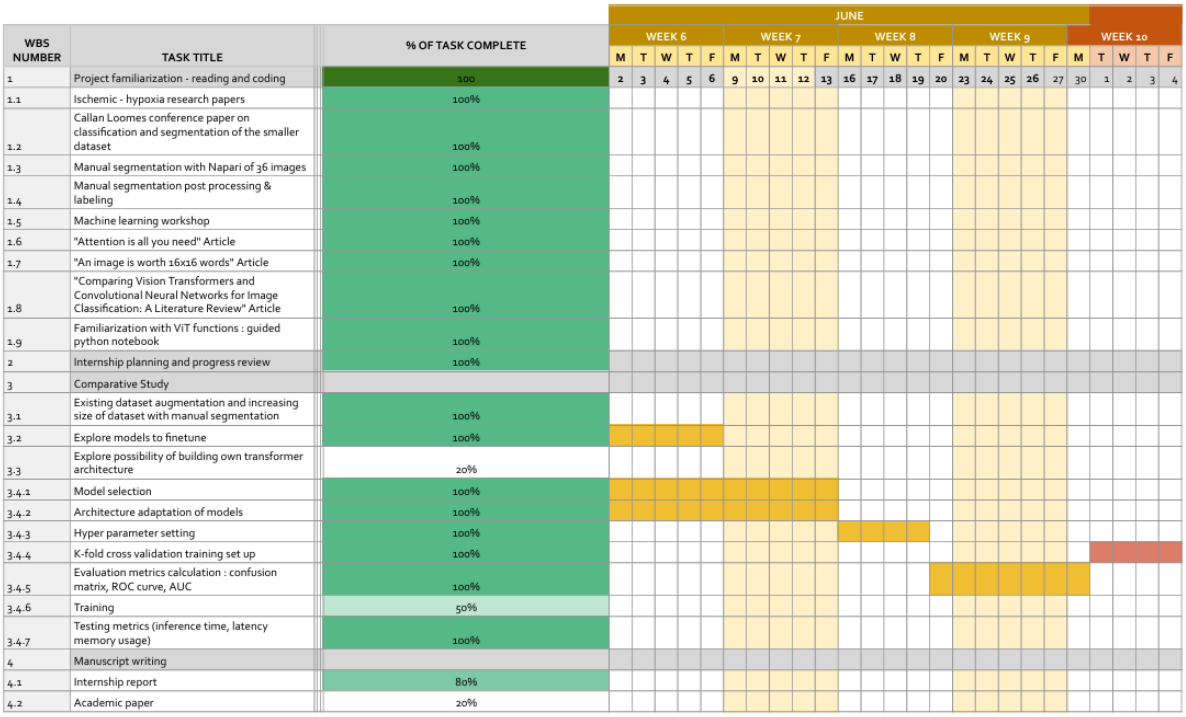
Annex 1 - Gantt chart

PROJECT TITLE	Dataset extension using Machine Learning for the segmentation and classification of hypoxic-ischemic neuronal cells in fetal sheep brain.	COMPANY NAME	Auckland Bio-engineering Insitute
PROJECT SUPERVISOR	ABBASI Hamid	DATE	04/2025 - 10/2025

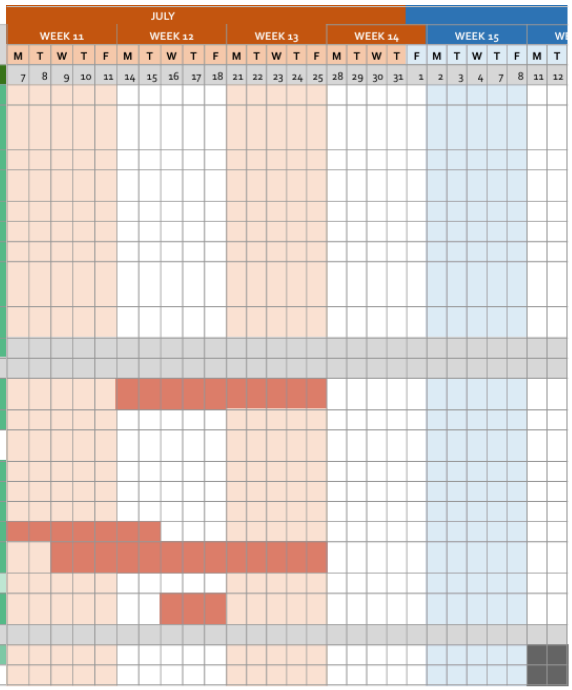
WBS NUMBER	TASK TITLE	% OF TASK COMPLETE
1	Project familiarization - reading and coding	100
1.1	Ischemic - hypoxia research papers	100%
1.2	Callan Loomes conference paper on classification and segmentation of the smaller dataset	100%
1.3	Manual segmentation with Napari of 36 images	100%
1.4	Manual segmentation post processing & labeling	100%
1.5	Machine learning workshop	100%
1.6	"Attention is all you need" Article	100%
1.7	"An image is worth 16x16 words" Article	100%
1.8	"Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review" Article	100%
1.9	Familiarization with ViT functions : guided python notebook	100%
2	Internship planning and progress review	100%
3	Comparative Study	
3.1	Existing dataset augmentation and increasing size of dataset with manual segmentation	100%
3.2	Explore models to finetune	100%
3.3	Explore possibility of building own transformer architecture	20%
3.4.1	Model selection	100%
3.4.2	Architecture adaptation of models	100%
3.4.3	Hyper parameter setting	100%
3.4.4	K-fold cross validation training set up	100%
3.4.5	Evaluation metrics calculation : confusion matrix, ROC curve, AUC	100%
3.4.6	Training	50%
3.4.7	Testing metrics (inference time, latency memory usage)	100%
4	Manuscript writing	
4.1	Internship report	80%
4.2	Academic paper	20%



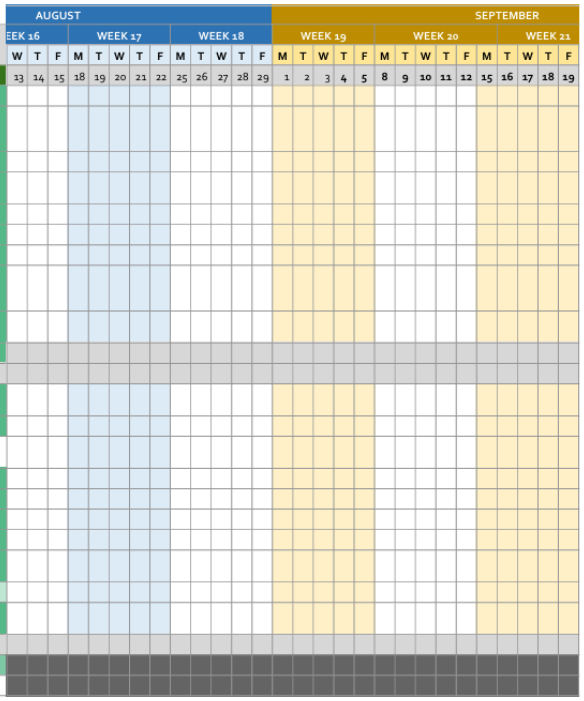
WBS NUMBER	TASK TITLE	% OF TASK COMPLETE
1	Project familiarization - reading and coding	100
1.1	Ischemic - hypoxia research papers	100%
1.2	Callan Loomes conference paper on classification and segmentation of the smaller dataset	100%
1.3	Manual segmentation with Napari of 36 images	100%
1.4	Manual segmentation post processing & labeling	100%
1.5	Machine learning workshop	100%
1.6	"Attention is all you need" Article	100%
1.7	"An image is worth 16x16 words" Article	100%
1.8	"Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review" Article	100%
1.9	Familiarization with ViT functions : guided python notebook	100%
2	Internship planning and progress review	100%
3	Comparative Study	
3.1	Existing dataset augmentation and increasing size of dataset with manual segmentation	100%
3.2	Explore models to finetune	100%
3.3	Explore possibility of building own transformer architecture	20%
3.4.1	Model selection	100%
3.4.2	Architecture adaptation of models	100%
3.4.3	Hyper parameter setting	100%
3.4.4	K-fold cross validation training set up	100%
3.4.5	Evaluation metrics calculation : confusion matrix, ROC curve, AUC	100%
3.4.6	Training	50%
3.4.7	Testing metrics (inference time, latency memory usage)	100%
4	Manuscript writing	
4.1	Internship report	80%
4.2	Academic paper	20%



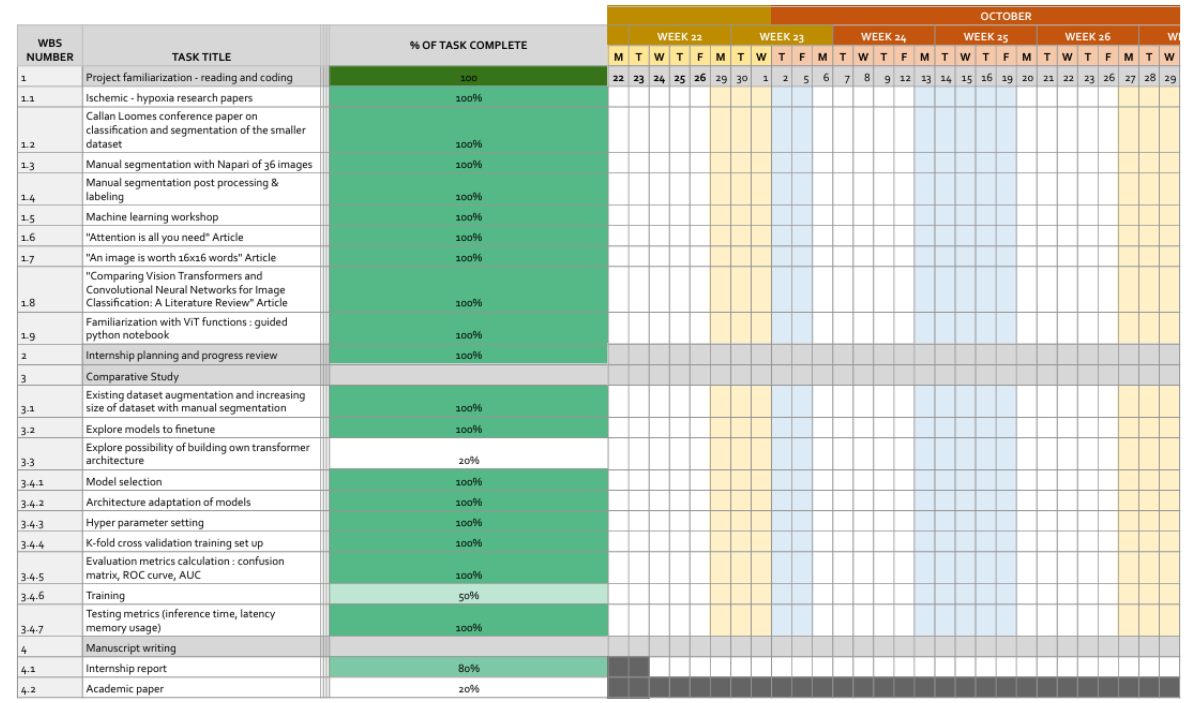
WBS NUMBER	TASK TITLE	% OF TASK COMPLETE
1	Project familiarization - reading and coding	100
1.1	Ischemic - hypoxia research papers	100%
1.2	Callan Loomes conference paper on classification and segmentation of the smaller dataset	100%
1.3	Manual segmentation with Napari of 36 images	100%
1.4	Manual segmentation post processing & labeling	100%
1.5	Machine learning workshop	100%
1.6	"Attention is all you need" Article	100%
1.7	"An image is worth 16x16 words" Article	100%
1.8	"Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review" Article	100%
1.9	Familiarization with ViT functions : guided python notebook	100%
2	Internship planning and progress review	100%
3	Comparative Study	
3.1	Existing dataset augmentation and increasing size of dataset with manual segmentation	100%
3.2	Explore models to finetune	100%
3.3	Explore possibility of building own transformer architecture	20%
3.4.1	Model selection	100%
3.4.2	Architecture adaptation of models	100%
3.4.3	Hyper parameter setting	100%
3.4.4	K-fold cross validation training set up	100%
3.4.5	Evaluation metrics calculation : confusion matrix, ROC curve, AUC	100%
3.4.6	Training	50%
3.4.7	Testing metrics (inference time, latency memory usage)	100%
4	Manuscript writing	
4.1	Internship report	80%
4.2	Academic paper	20%



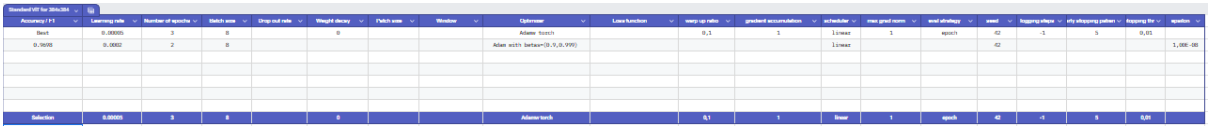
WBS NUMBER	TASK TITLE	% OF TASK COMPLETE
1	Project familiarization - reading and coding	100
1.1	Ischemic - hypoxia research papers	100%
1.2	Callan Loomes conference paper on classification and segmentation of the smaller dataset	100%
1.3	Manual segmentation with Napari of 36 images	100%
1.4	Manual segmentation post processing & labeling	100%
1.5	Machine learning workshop	100%
1.6	"Attention is all you need" Article	100%
1.7	"An image is worth 16x16 words" Article	100%
1.8	"Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review" Article	100%
1.9	Familiarization with ViT functions : guided python notebook	100%
2	Internship planning and progress review	100%
3	Comparative Study	
3.1	Existing dataset augmentation and increasing size of dataset with manual segmentation	100%
3.2	Explore models to finetune	100%
3.3	Explore possibility of building own transformer architecture	20%
3.4.1	Model selection	100%
3.4.2	Architecture adaptation of models	100%
3.4.3	Hyper parameter setting	100%
3.4.4	K-fold cross validation training set up	100%
3.4.5	Evaluation metrics calculation : confusion matrix, ROC curve, AUC	100%
3.4.6	Training	50%
3.4.7	Testing metrics (inference time, latency memory usage)	100%
4	Manuscript writing	
4.1	Internship report	80%
4.2	Academic paper	20%



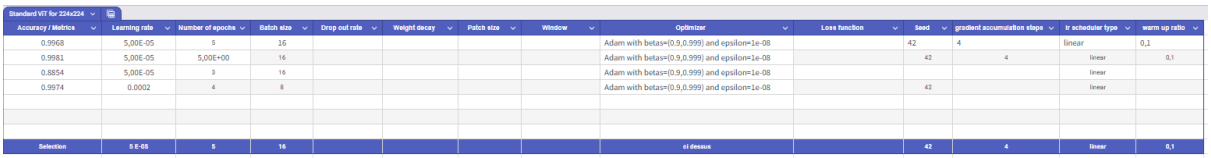
WBS NUMBER	TASK TITLE	% OF TASK COMPLETE
1	Project familiarization - reading and coding	100
1.1	Ischemic - hypoxia research papers	100%
1.2	Callan Loomes conference paper on classification and segmentation of the smaller dataset	100%
1.3	Manual segmentation with Napari of 36 images	100%
1.4	Manual segmentation post processing & labeling	100%
1.5	Machine learning workshop	100%
1.6	"Attention is all you need" Article	100%
1.7	"An image is worth 16x16 words" Article	100%
1.8	"Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review" Article	100%
1.9	Familiarization with ViT functions : guided python notebook	100%
2	Internship planning and progress review	100%
3	Comparative Study	
3.1	Existing dataset augmentation and increasing size of dataset with manual segmentation	100%
3.2	Explore models to finetune	100%
3.3	Explore possibility of building own transformer architecture	20%
3.4.1	Model selection	100%
3.4.2	Architecture adaptation of models	100%
3.4.3	Hyper parameter setting	100%
3.4.4	K-fold cross validation training set up	100%
3.4.5	Evaluation metrics calculation : confusion matrix, ROC curve, AUC	100%
3.4.6	Training	50%
3.4.7	Testing metrics (inference time, latency memory usage)	100%
4	Manuscript writing	
4.1	Internship report	80%
4.2	Academic paper	20%



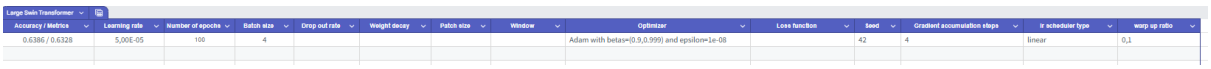
Annex 2 - Hyperparameters tables



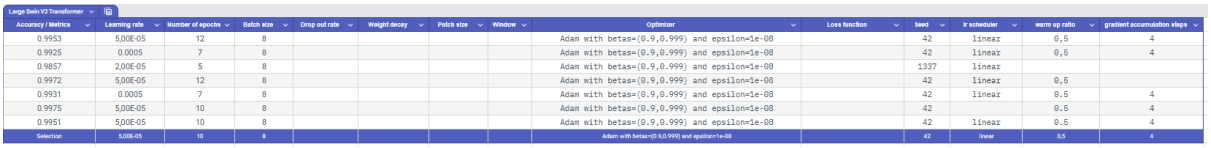
Accuracy / F1	Learning rate	Number of epochs	Batch size	Drop out rate	Weight decay	Patch size	Window	Optimizer	Loss function	warp up ratio	gradient accumulation	scheduler	max grad norm	eval strategy	seed	logging steps	stty stopping patien	logging thr	epsilon
Best	0.00005	3	8		0			Adamw_torch		0,1	1	linear	1	epoch	42	-1	5	0,01	
0.9695	0.0002	2	8					Adam with betas=(0.9,0.999)				linear			42				1,00E-08
Selection	0.00005	3	8		0			Adamw torch		0,1	1	linear	1	epoch	42	-1	5	0,01	



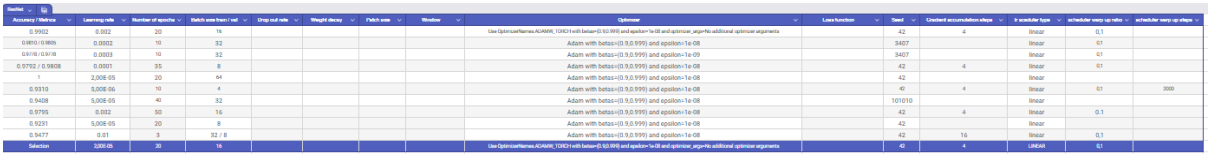
Accuracy / Metrics	Learning rate	Number of epochs	Batch size	Drop out rate	Weight decay	Patch size	Window	Optimizer	Loss function	Seed	gradient accumulation steps	lr scheduler type	warm up ratio
0.9968	5,00E-05	5	16					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	4	linear	0,1
0.9981	5,00E-05	5,00E+00	16					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	4	linear	0,1
0.8854	5,00E-05	3	16					Adam with betas=(0.9,0.999) and epsilon=1e-08				linear	
0.9974	0.0002	4	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42		linear	
Selection	5 E-05	5	16					ci dessus		42	4	linear	0,1



Accuracy / Metrics	Learning rate	Number of epochs	Batch size	Drop out rate	Weight decay	Patch size	Window	Optimizer	Loss function	Seed	Gradient accumulation steps	lr scheduler type	warp up ratio
0.6386 / 0.6328	5,00E-05	100	4					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	4	linear	0,1



Accuracy / Metrics	Learning rate	Number of epochs	Batch size	Drop out rate	Weight decay	Patch size	Window	Optimizer	Loss function	Seed	lr scheduler	warm up ratio	gradient accumulation steps
0.9953	5,00E-05	12	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	linear	0,5	4
0.9925	0.0005	7	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	linear	0,5	4
0.9857	2,00E-05	5	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		1337	linear		
0.9972	5,00E-05	12	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	linear	0,5	
0.9931	0.0005	7	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	linear	0.5	4
0.9975	5,00E-05	10	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42		0.5	4
0.9951	5,00E-05	10	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	linear	0.5	4
Selection	5,00E-05	10	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	linear	0,5	4



Accuracy / Metrics	Learning rate	Number of epochs	Batch size train / val	Drop out rate	Weight decay	Patch size	Window	Optimizer	Loss function	Seed	Gradient accumulation steps	lr scheduler type	scheduler warp up ratio	scheduler warp up steps
0.9902	0.002	20	16					Use OptimizerNames.ADAMW_TORCH with betas=(0.92,0.999) and epsilon=1e-08 and optimizer_args=No additional optimizer arguments		42	4	linear	0,1	
0.9810 / 0.9826	0.0002	10	32					Adam with betas=(0.9,0.999) and epsilon=1e-08		3407		linear	0,1	
0.9718 / 0.9718	0.0003	10	32					Adam with betas=(0.9,0.999) and epsilon=1e-09		3407		linear	0,1	
0.9792 / 0.9808	0.0001	35	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	4	linear	0,1	
1	2,00E-05	20	64					Adam with betas=(0.9,0.999) and epsilon=1e-08		42		linear		
0.9310	5,00E-06	10	4					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	4	linear	0,1	3000
0.9408	5,00E-05	40	32					Adam with betas=(0.9,0.999) and epsilon=1e-08		101010		linear		
0.9795	0.002	50	16					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	4	linear	0,1	
0.9231	5,00E-05	20	8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42		linear		
0.9477	0.01	3	32 / 8					Adam with betas=(0.9,0.999) and epsilon=1e-08		42	16	linear	0,1	
Selection	2,00E-05	20	16					Use OptimizerNames.ADAMW_TORCH with betas=(0.92,0.999) and epsilon=1e-08 and optimizer_args=No additional optimizer arguments		42	4	LINEAR	0,1	

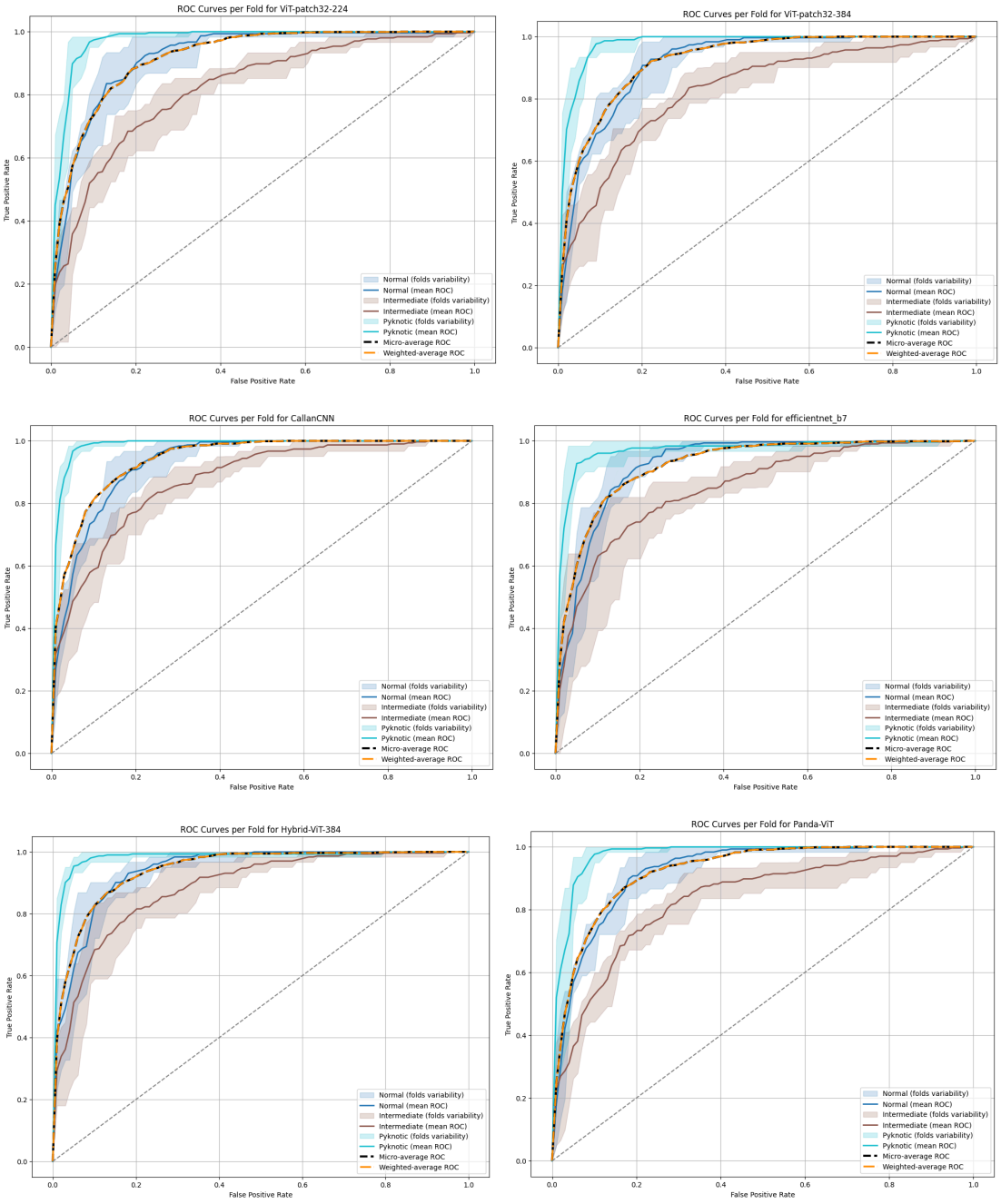
EfficientNet

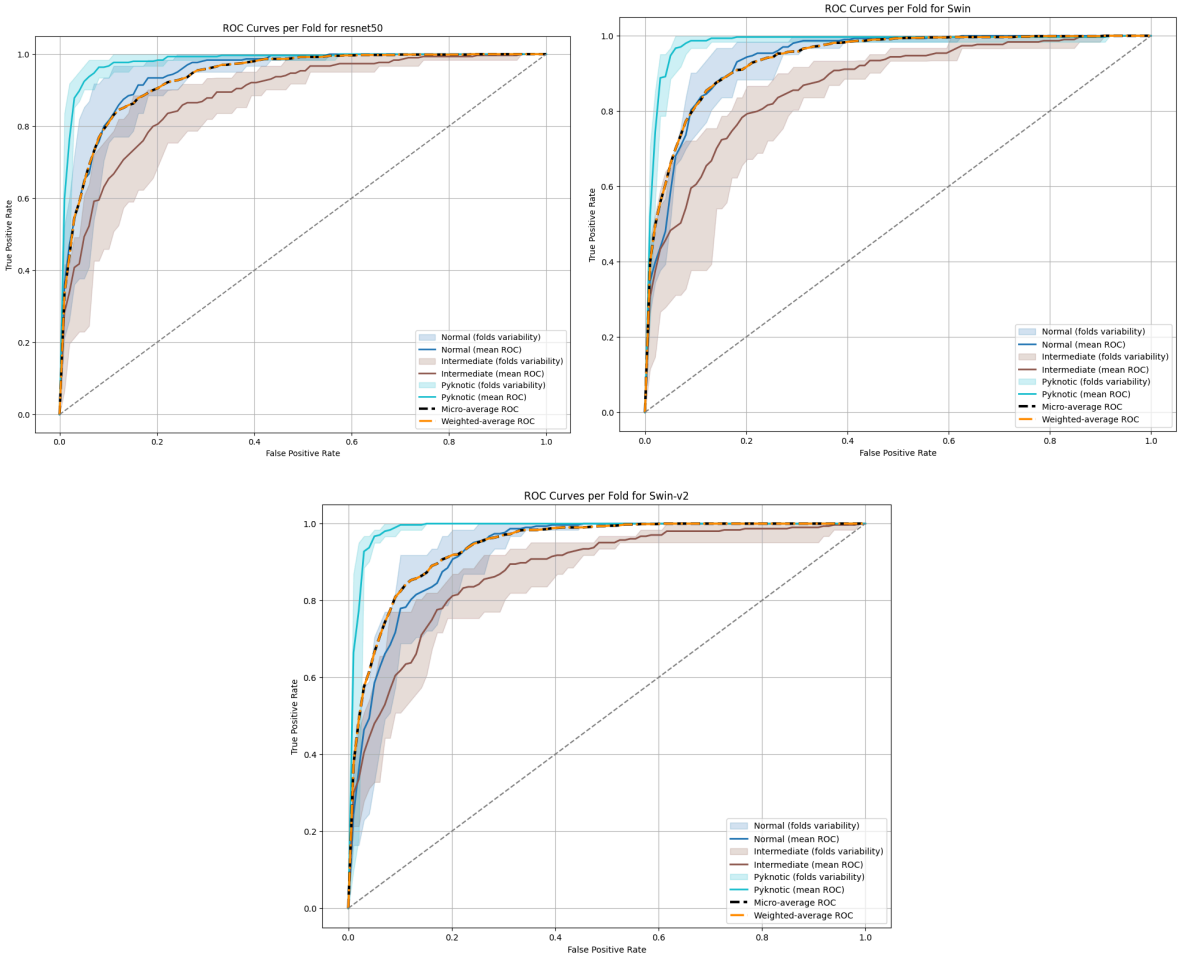
Accuracy / Metrics	Learning rate	Number of epochs	Batch size	Drop out rate	Weight decay	Patch size	Window	Optimizer	Loss function	Seed	lr scheduler type	gradient accumulation steps	lr scheduler warp up ratio	lr scheduler warp up steps	Mixed precision training
0.9097	2,00E-05	100	8					Use OptimizerNames.ADA		1337	linear	4			
0.9253	5,00E-05	5	32					Use OptimizerNames.ADA		42	linear		0,1		
99.04	5,00E-04	25	32					AdamW							
0.9128	5,00E-05	30	64					Use OptimizerNames.ADA		42	cosine with restarts			256	Native AMP
	5,00E-04	25	32					0.9,0.999) and epsilon=1e-08 an		42	linear	4	0,1	256	Native AMP

Hybrid ViT

Accuracy / Metrics	Learning rate	Number of epochs	Batch size	Drop out rate	Weight decay	Patch size	Window	Optimizer	Loss function	Seed	Gradient accumulation steps	lr scheduler type	warp up ratio	Mixed precision training	Warm up steps
0.9942	5,00E-05	3	8					ss=(0.9,0.999) and		42	4	linear	0.1	Native AMP	
0.9286	5,00E-05	15	32					and epsilon=1e-08		42		ine_with_resta	0.1		
0.9195	0.0003	30	64					and epsilon=1e-08		42		ine_with_resta		Native AMP	256
0.9664	5,00E-05	30	32					and epsilon=1e-08		42		ine_with_resta		Native AMP	512
0.9430	5,00E-05	15	32					and epsilon=1e-08		42		ine_with_resta		Native AMP	512
	0,00005	15	32					0.9,0.999) and epsilon=1e-08 an		42	4	linear	0,1	Native AMP	

Annex 3 - ROC Curves for all models





Annex 4 -Training time

Annex 6 - Ethical approval for the experiments

The University of Auckland
Private Bag 92019
Auckland, New Zealand

Level 3, 49 Symonds Street
Telephone: 64 9 373 7599
Extension: 86356

UNIVERSITY OF AUCKLAND ANIMAL ETHICS COMMITTEE (AEC)

14/05/2025

MEMORANDUM TO:

Dr Joanne Davidson
Physiology

Application for ethics approval (Our Ref. AEC28175): Research Application approved

The Committee considered the animal ethics application for your project entitled "Pathogenesis, detection and treatment of perinatal brain and organ injury".

The Committee is pleased to advise you that this application has now been approved for a period of three years.

The approval date is **14/05/2025.**

The expiry date is **14/05/2028.**

Conditions of approval

- All deaths that occur prior to the planned end of experiments must be notified to the Animal Welfare Officer so that a post-mortem may be performed, if considered necessary. This includes all animals that are found dead or need to be euthanised due to any health or welfare compromise or abnormalities that make them not fit for purpose.
- Please note the requirement of reporting animal use under the Animal Welfare Act 1999. As Principal Investigator, it is your statutory responsibility to provide to this office:

1. An Animal Usage Return (AUR) for incorporation into the University's consolidated return to MPI. This must occur within 3 months of the end of the approval (or end of the research, if prior to that). If the approval ends in November or December of a certain year, the information must be provided by 31 January of the following year.
2. An End of Approval Report at completion of the project. This must occur within 3 months, or it may impact your ability to apply for future AEC approvals.

IDAOs for Drugs used in the Management of Animals:

This approval does not extend to any IDAOs associated with this application. Please contact the Animal Welfare Officer for assistance with IDAOs, even if they were submitted with your application. You are required to have authorised copies of these IDAOs before starting any animal work requiring these drugs.

Requesting animals from the VJU:

If you are obtaining animals from the VJU, or using the VJU to house your animals, your AEC approval will be automatically loaded into the Tick@lab database and be available to you the next day. If this does not occur then please email the VJU team at vjuorders@auckland.ac.nz

All required forms, general information on the animal ethics procedures, and information on training can be found on the Staff Intranet or can be provided by the Animal Ethics Administrator on request.

If you have any queries regarding your ethics application or wish to discuss general matters relating to ethics approvals, please contact the Ethics & Regulatory Coordinator in the first instance at animalethics@auckland.ac.nz

Please quote reference **AEC28175** for all communication with the AEC regarding this application.

Animal Ethics Administrator

University of Auckland Animal Ethics Committee