Evaluating Acceleration Techniques for Candidate Evaluation in Genetic Neural Architecture Search

Foteini Dervisi

Supervisor: Professor Konstantinos Margaritis

Neural Architecture Search

- Automating the design of artificial neural network architectures
- The success of deep learning techniques relies on making appropriate neural architecture choices for the task at hand
- Manual architecture design is a time-consuming and error-prone process

Components

- **Search space**: defines which neural architectures are eligible for selection by the algorithm
- **Optimisation method**: used for the exploration of the chosen search space (e.g. evolutionary algorithms, reinforcement learning)
- **Candidate evaluation method**: estimates the quality of candidate neural network architectures

Benchmarks

- Benchmarks facilitate the comparison between different algorithms by providing standardised environments for the evaluation of different algorithms.
- **NAS-Bench-101**: contains CNN architectures that have been trained and evaluated on CIFAR-10
- **NAS-Bench-201/NATS-Bench**: contains CNN architectures that have been trained and evaluated on CIFAR-10, CIFAR-100 and ImageNet-16-120

Genetic Algorithm

- Evolutionary algorithms: population-based metaheuristics inspired by the biological process of natural selection
- **Iterative procedure** to discover high-performing neural architectures
- **Parents** are defined by using tournament selection
- **Offsprings** are generated from parents with the use of bitwise mutation
- Use of a **fitness score** for the evaluation of architectures

NAS-EA-FA V2

- Genetic algorithm
- **Fitness approximation** (XGBoost, Graph Convolutional Network) to accelerate the search
- **Data augmentation** and **enhanced diversity**: efficiency and stability
- Individuals with **top *K* predicted fitness** and **top *H* distance from the previously evaluated architectures** are trained and evaluated in each iteration

C. Pan and X. Yao, "Neural Architecture Search Based on Evolutionary Algorithms with Fitness Approximation," *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1-8, doi: 10.1109/IJCNN52387.2021.9533986.

NAS Without Training

- NAS algorithms tend to be slow and expensive
- **Partial prediction** of a network's trained accuracy from its **initial state**
- Works for neural networks with **ReLU activations**
- Overlap of activations between datapoints in untrained networks
- Evaluation of architectures without any training **in a matter of seconds** on a single GPU

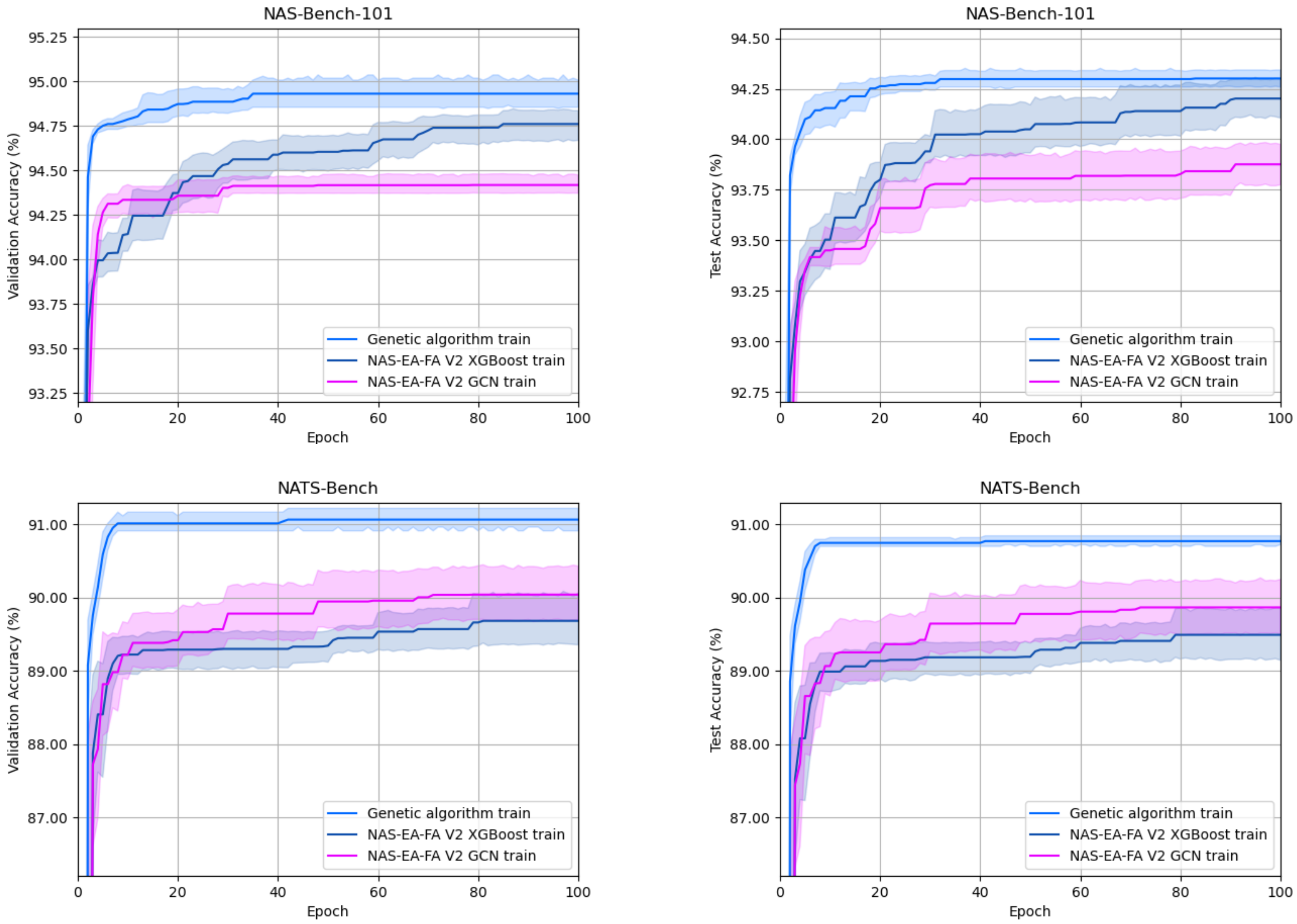
J. Mellor, J. Turner, A. Storkey and E. J. Crowley, "Neural Architecture Search without Training," *2021 International Conference on Machine Learning (ICML)*, 2021

Experiments

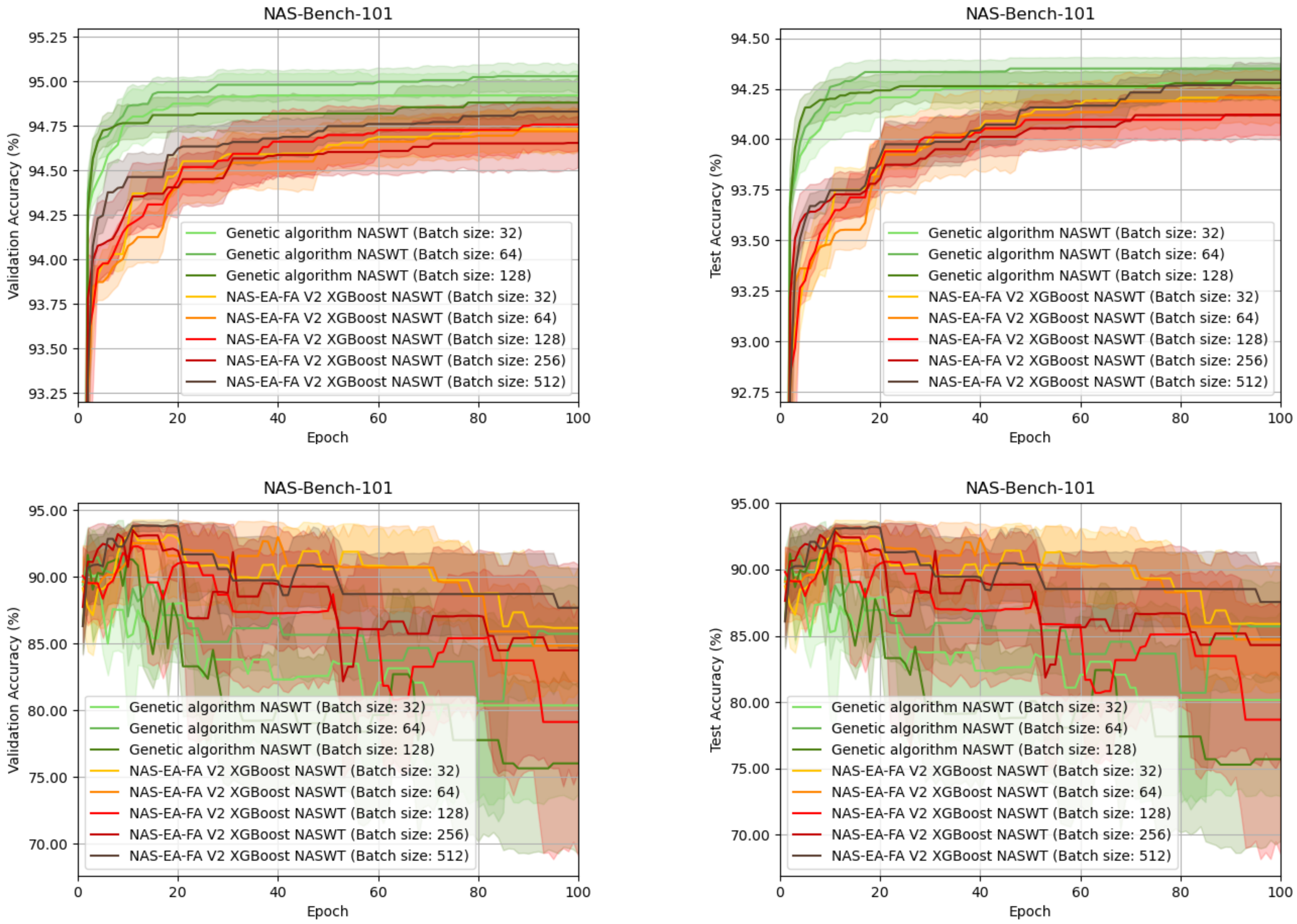
- **Genetic algorithm** with **validation accuracy** as fitness
- **Genetic algorithm** with **NASWT score** as fitness
- **NAS-EA-FA V2 (XGBoost)** with **validation accuracy** as fitness
- **NAS-EA-FA V2 (GCN)** with **validation accuracy** as fitness
- **NAS-EA-FA V2 (XGBoost)** with **NASWT score** as fitness

**All experiments were performed on the NAS-Bench-101 dataset and on the NATS-Bench topology search space dataset.*

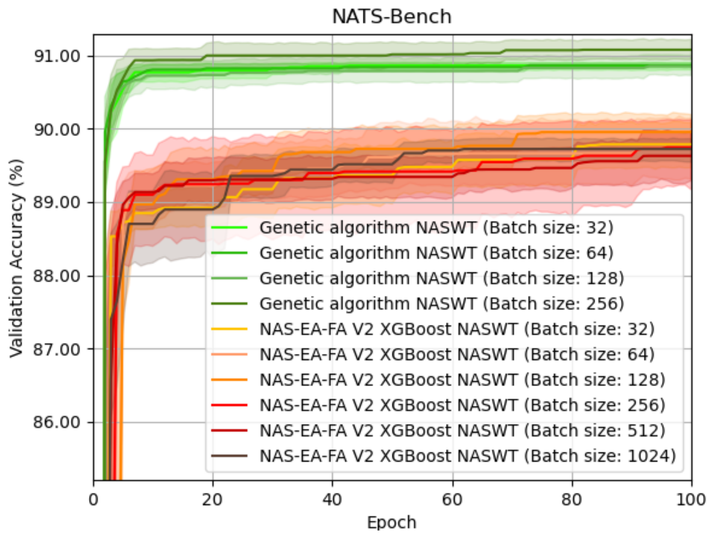
Results – Performance Statistics

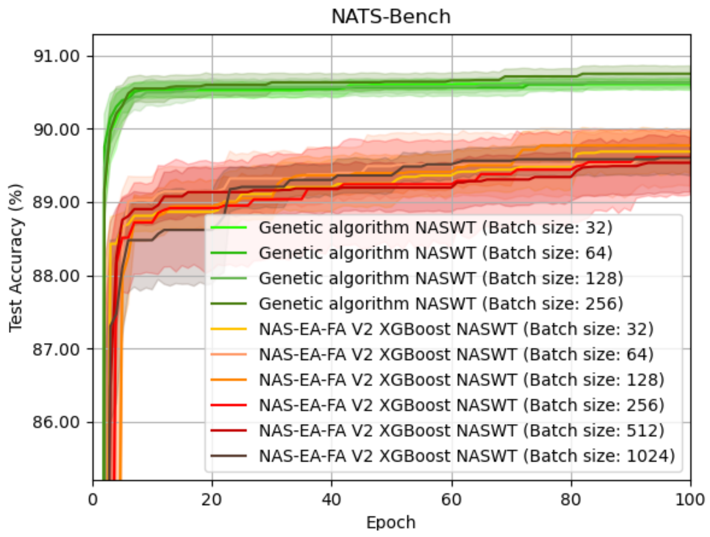


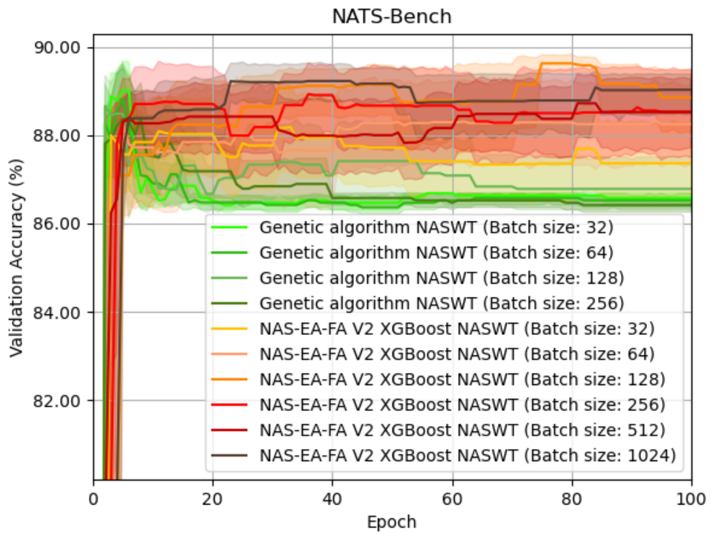
Results – Performance Statistics

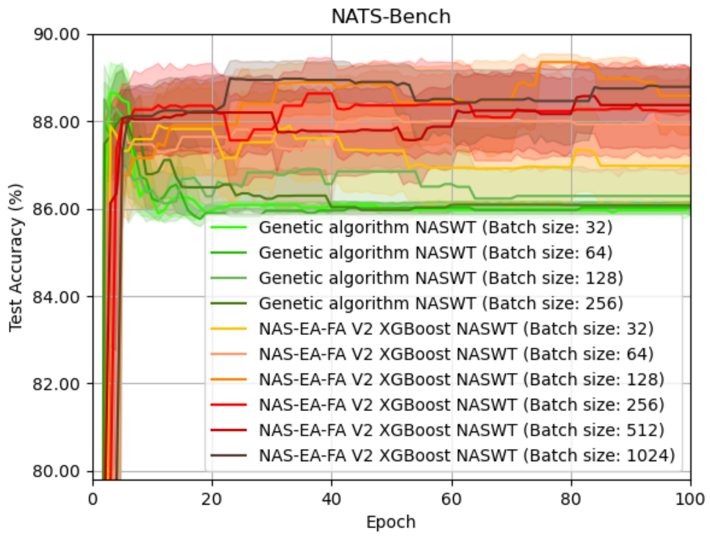


Results – Performance Statistics

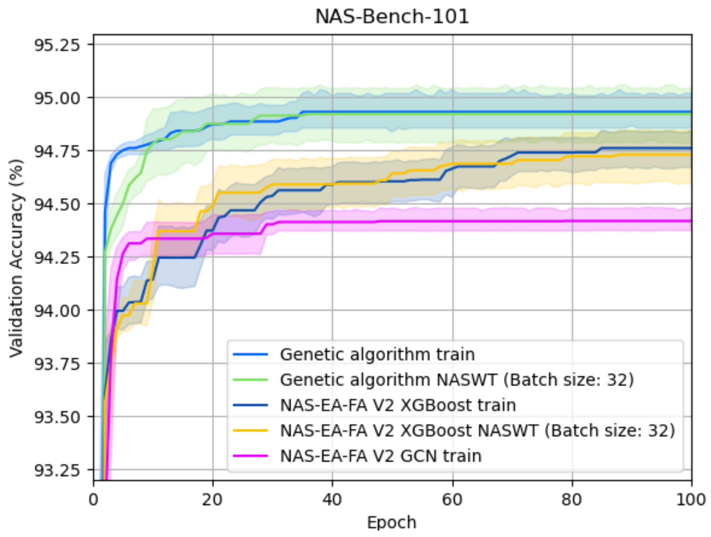


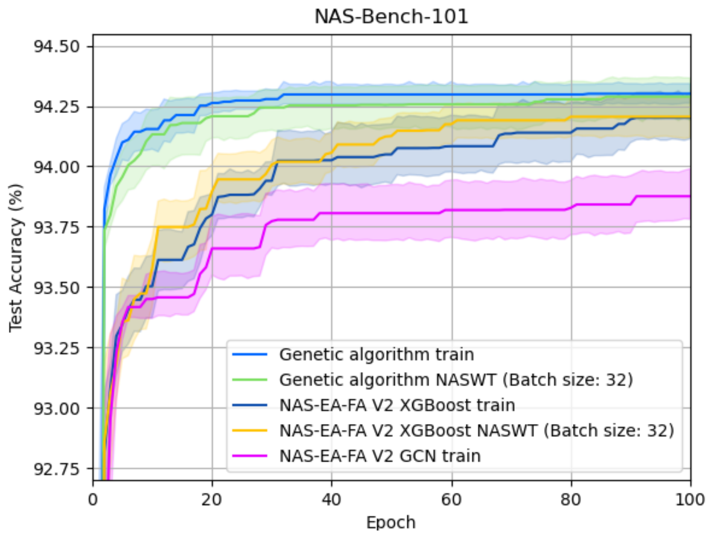


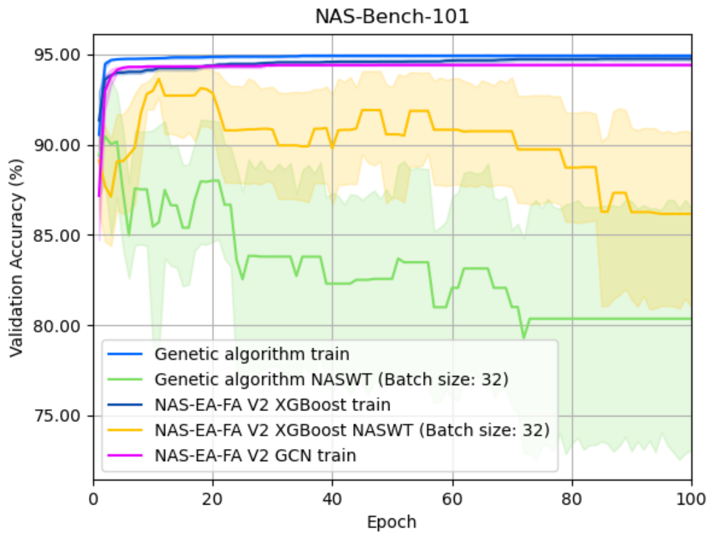


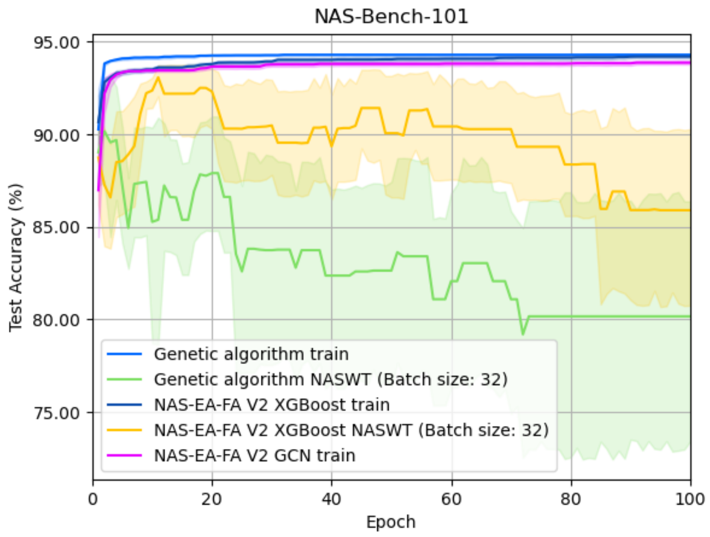


Results – Performance Statistics

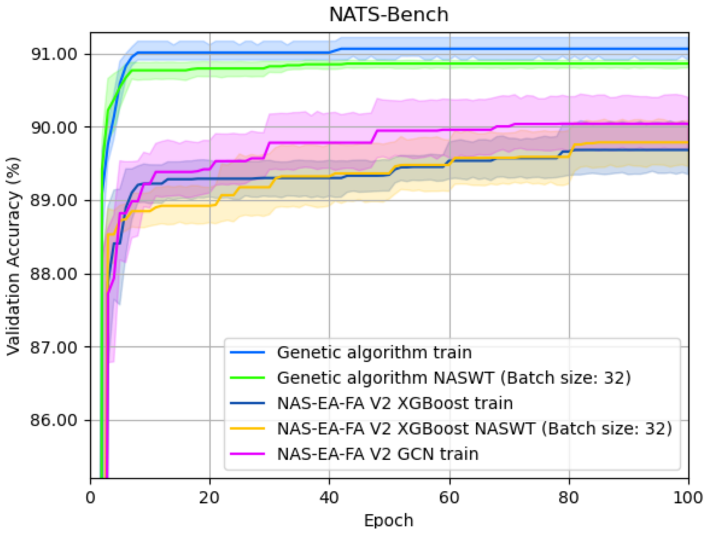


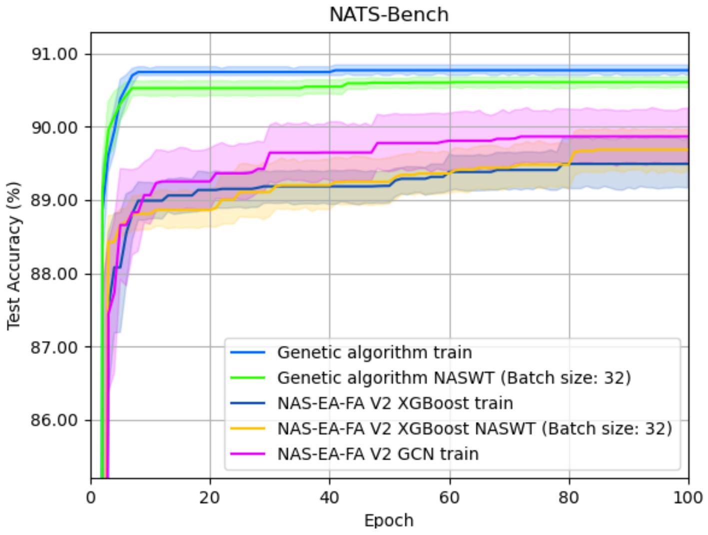


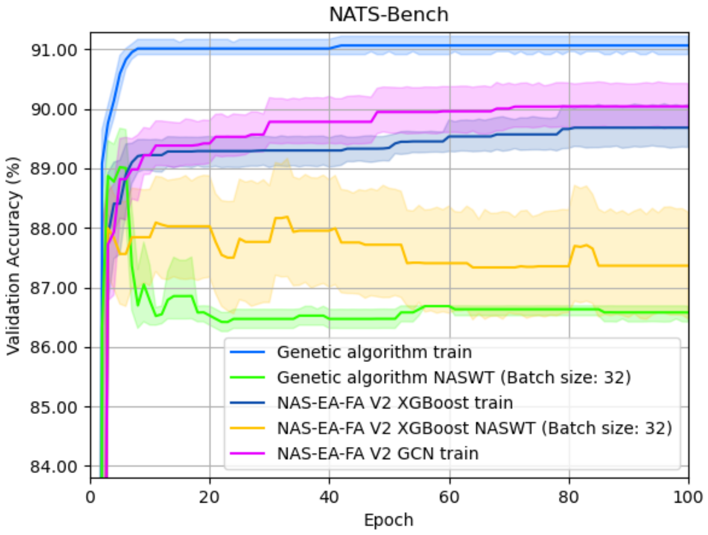


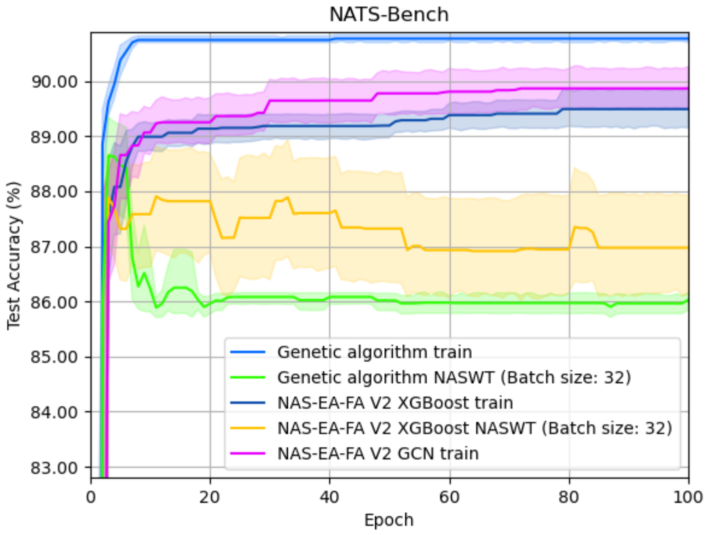


Results – Performance Statistics

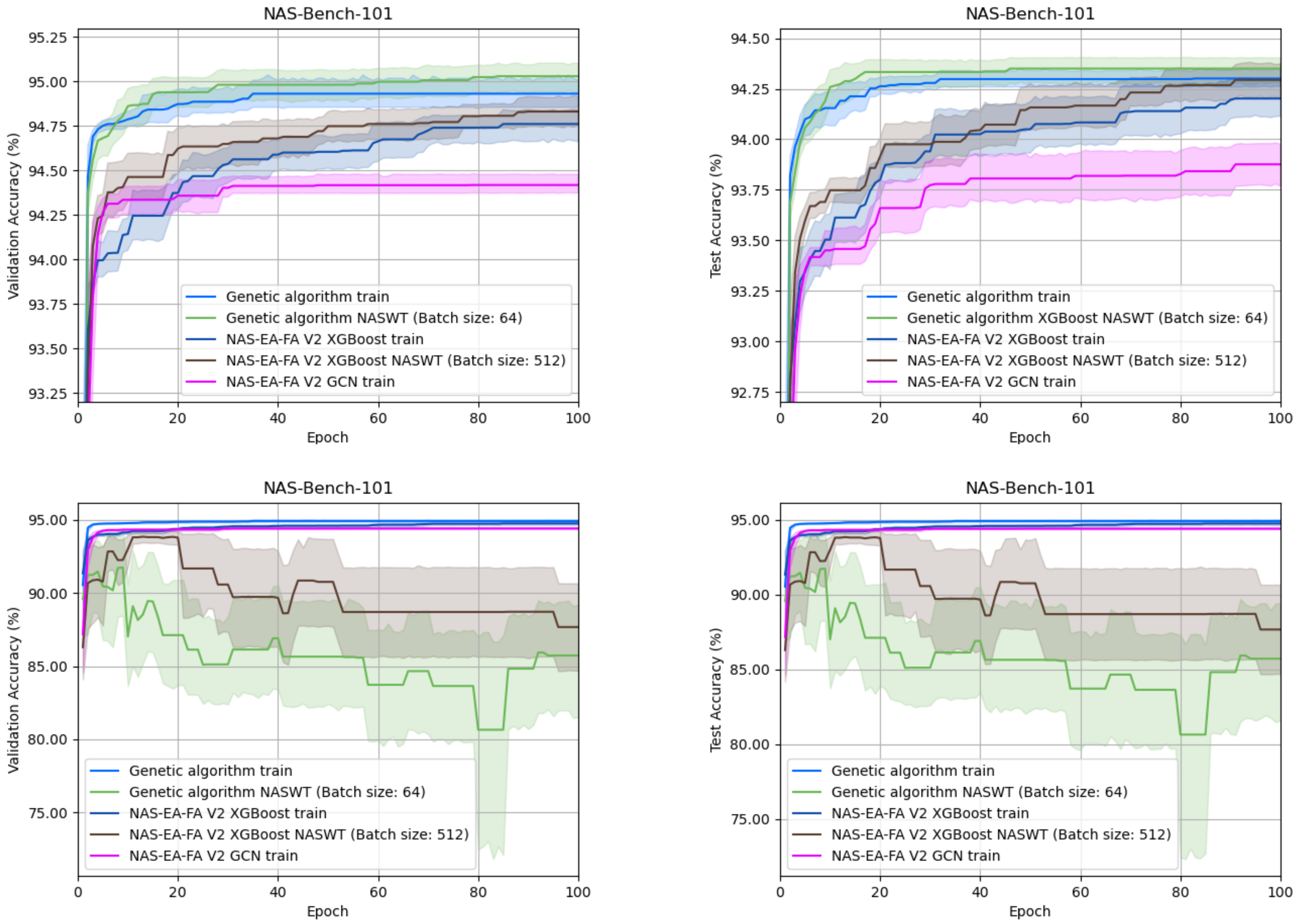




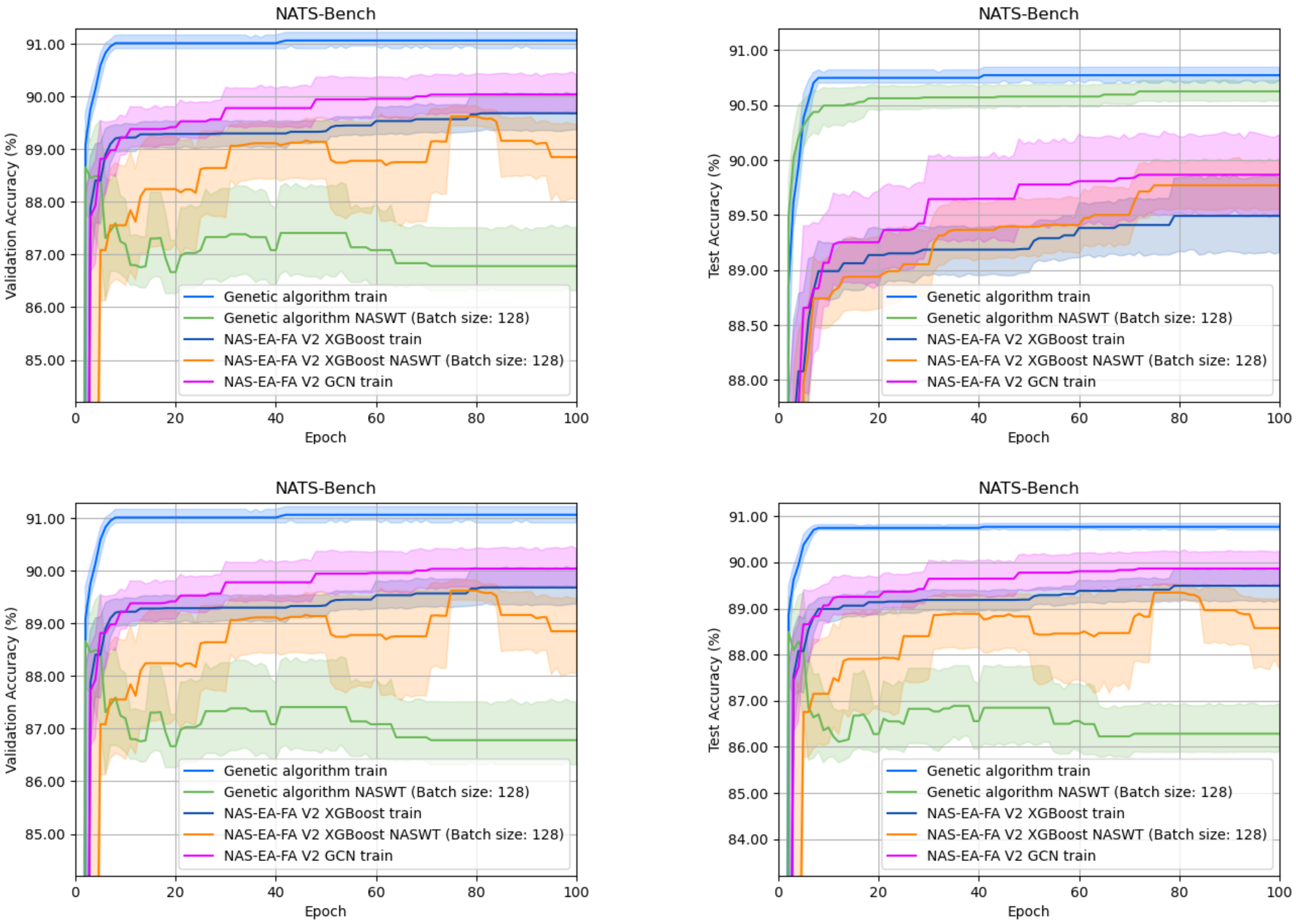




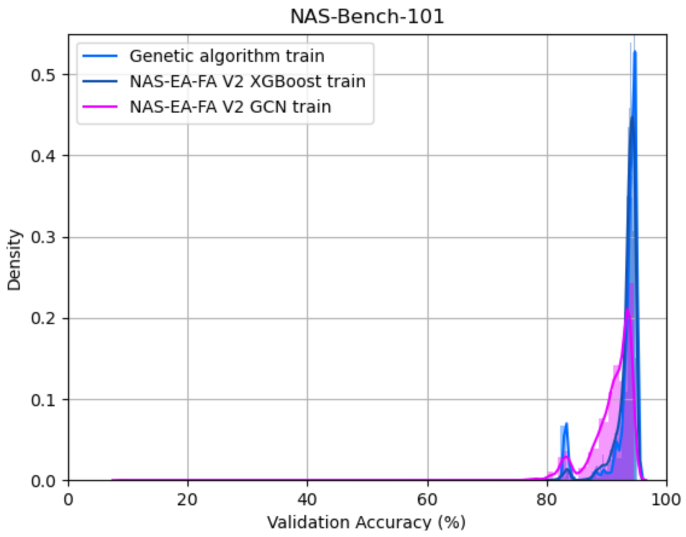
Results – Performance Statistics

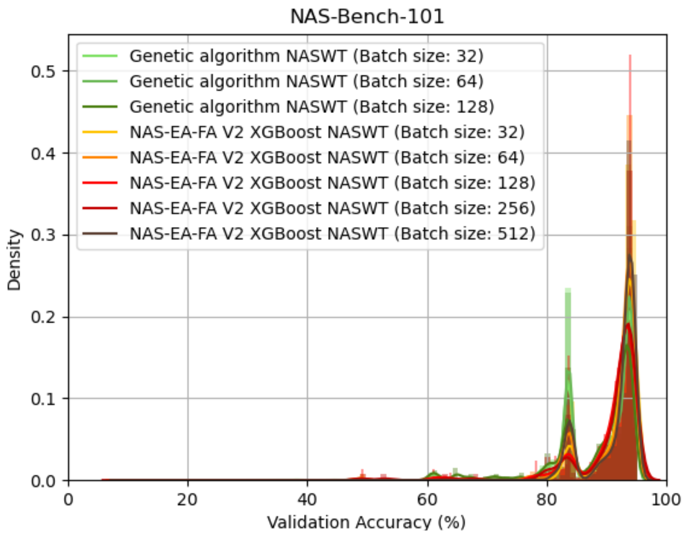


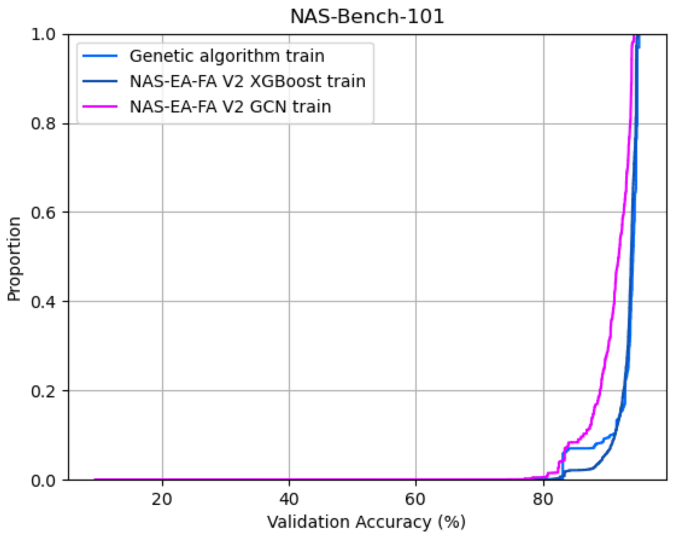
Results – Performance Statistics

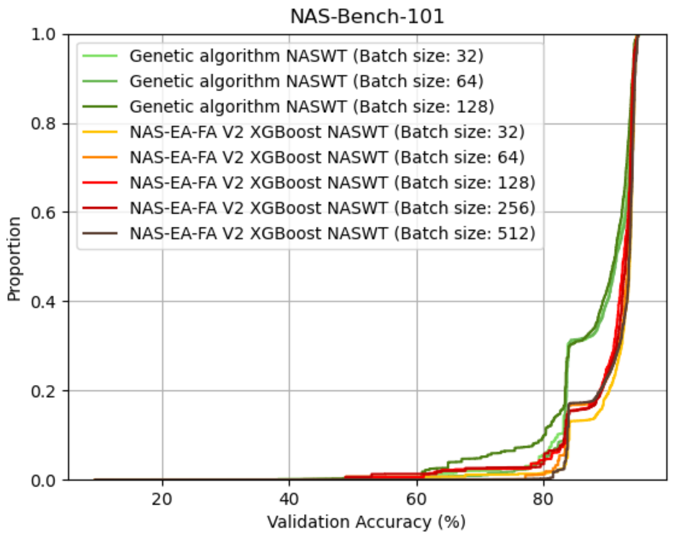


Results – Distributions

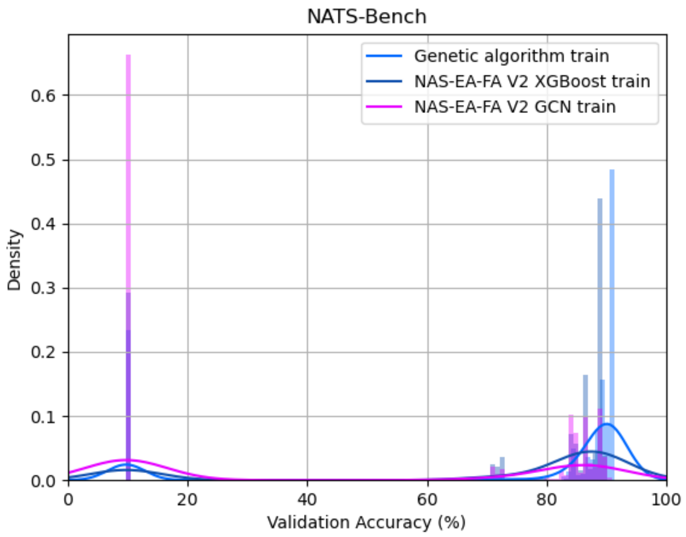


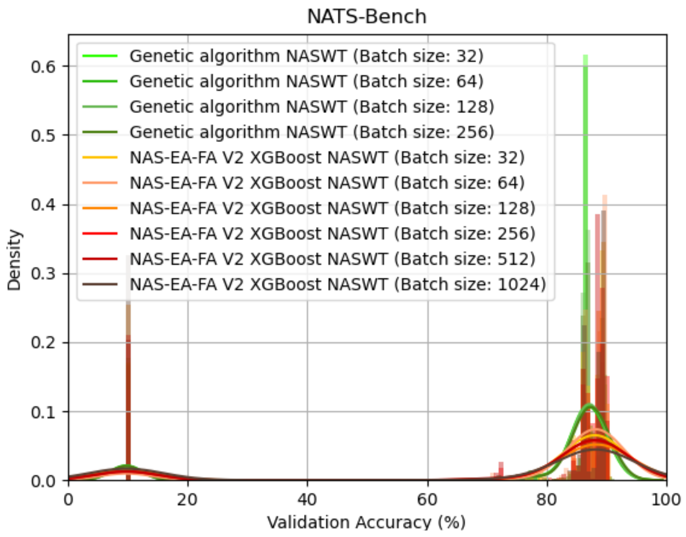


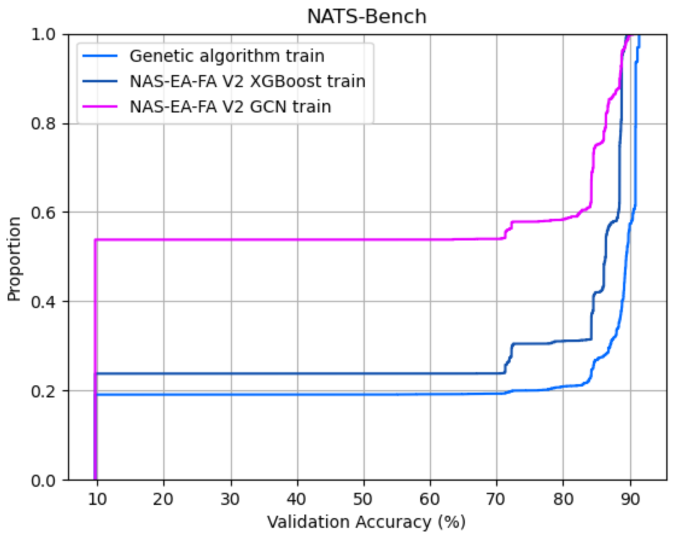


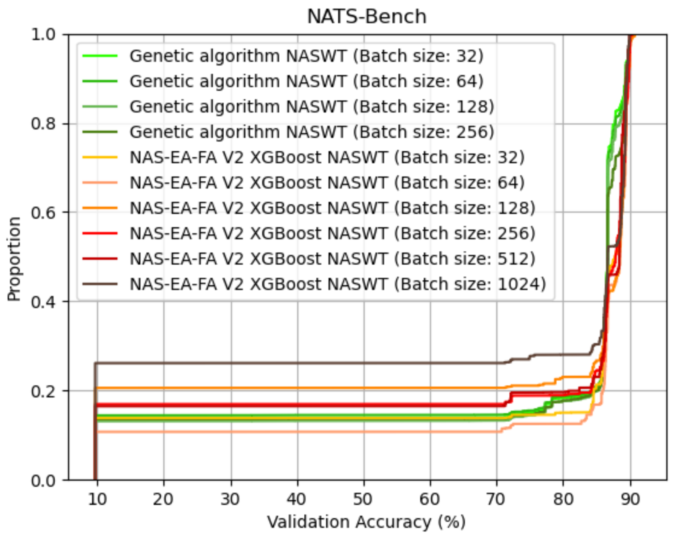


Results – Distributions









Conclusions

- **NAS-EA-FA V2 narrows down the search space** due to the bias introduced by the process that is used for the selection of training data for the fitness approximation model
- The **NASWT score** succeeds in guiding the search to find high-performing architectures but **cannot distinguish them** from worse-performing architectures
- **Increasing the size of the batch** used for the calculation of the NASWT score **does not always lead to improved performance**

Conclusions

- The **simple genetic algorithm** and **NAS-EA-FA V2** produce architectures with validation accuracies that **belong to different distributions** regardless of the chosen fitness score
- **NAS-EA-FA V2** with the use of the **validation accuracy as fitness score** is the best choice for the acceleration of the NAS procedure

Future Work

- Examine the behaviour of **other fitness approximation models**
- Examine the use of **other metaheuristic optimisation methods** (e.g. particle swarm optimisation, ant colony optimisation)
- Investigate the use of acceleration techniques in cases where **other optimisation algorithms** are used (e.g. reinforcement learning, Bayesian optimisation)

Thank you!
Questions?