

1 **Application of ConvNeXt with Transfer Learning and Data Augmentation**
2 **for Malaria Parasite Detection in Resource-Limited Settings Using**
3 **Microscopic Images**

4

5 Outlwile Pako Mmileng^{1*}, Albert Whata², Micheal Olusanya³, Siyabonga Mhlongo⁴

6

7 1 Centre for Applied Data Science, University of Johannesburg, Johannesburg, South Africa

8 2 Department of Statistics, University of Pretoria, Pretoria, South Africa

9 3 Department of Computer Science and Information Technology, Sol Plaatje University,

10 Kimberley, South Africa

11 4 Department of Applied Information Systems, University of Johannesburg, Johannesburg,

12 South Africa

13

14 * Corresponding author

15 Email: 223257123@student.uj.ac.za (OPM)

16

17 **Abstract**

18

19 Malaria is one of the most widespread and deadly diseases across the globe, especially in sub-
20 Saharan Africa and other parts of the developing world. This is primarily because of incorrect
21 or late diagnosis. Existing diagnostic techniques mainly depend on the microscopic
22 identification of parasites in the blood smear stained with special dyes, which have drawbacks
23 such as being time-consuming, depending on skilled personnel and being vulnerable to errors.

24

25 This work seeks to overcome these challenges by proposing a deep learning-based solution in
26 the ConvNeXt architecture incorporating transfer learning and data augmentation to automate
27 malaria parasite identification in thin blood smear images. This study's dataset was a set of
28 blood smear images of equal numbers of parasitised and uninfected samples drawn from a
29 public database of malaria patients in Bangladesh. To detect malaria in the given dataset of
30 parasitised and uninfected blood smears, the ConvNeXt models were fine-tuned. To improve
31 the effectiveness of these models, a vast number of data augmentation strategies was used so
32 that the models could work well in various image capture conditions and perform well even in
33 environments with limited resources. The ConvNeXt Tiny model performed better, particularly
34 the re-tuned version, than other models, such as Swin Tiny, ResNet18, and ResNet50, with an
35 accuracy of 95%. On the other hand, the re-modified version of the ConvNeXt V2 Tiny model
36 reached 98% accuracy. These findings show the potential to implement ConvNeXt-based
37 systems in regions with scarce healthcare facilities for effective and affordable malaria
38 diagnosis.

39

40 Keywords: ConvNeXt, deep learning, malaria detection, medical imaging, resource-limited
41 settings transfer learning,

42

43 1. Introduction

44

45 According to the World Health Organization (WHO), there were 241 million cases and 627,000
46 deaths due to malaria in 2020, with the disease primarily affecting low-income countries (1).
47 Malaria is an infectious disease caused by parasitic protozoa of the genus *plasmodium*, of which
48 *plasmodium falciparum* and *plasmodium vivax* are the most pathogenic to man (2). These are
49 transmitted through the bites of infected *Anopheles* mosquitoes. In malaria-endemic areas,
50 timely and correct diagnosis is critical to prevent complications and minimise transmission.
51 However, access to accurate diagnostic instruments is still problematic, especially in low-
52 income regions.

53

54 Malaria diagnosis is usually performed by examination of thick and thin giemsa stained blood
55 films where the laboratory technologists manually use microscopes to look for the malaria
56 parasites (3, 4). Even though this approach is reasonably practical, it is rather time-consuming,
57 somewhat subjective and dependent on qualified specialists. Some reasons include technician
58 fatigue, poor imaging conditions, or variability in the blood smear preparation. Using artificial
59 intelligence, specifically deep learning, to develop automated diagnostic tools can be an
60 excellent solution to increase diagnostic efficiency and decrease the workload of healthcare
61 professionals in limited resource settings (5).

62

63 Convolutional neural networks (CNNs), a type of deep learning, are very effective in the
64 automated diagnosis of medical images. CNNs have been applied in many applications,
65 including disease diagnosis, object identification, and segmentation (6). However, standard
66 CNNs are data-hungry and labelled medical data are limited in many regions worldwide (7).
67 To address these problems, this study uses the ConvNeXt architecture, which combines the

68 ease of use of conventional CNNs with the hierarchical feature extraction of the latest models,
69 such as vision transformers (ViTs).

70

71 The objectives of this study are as follows:

- 72 • To build a robust automated malaria diagnostic tool using the ConvNeXt architecture.
- 73 • To improve performance using transfer learning from pre-trained models on large
74 datasets such as ImageNet.
- 75 • To investigate how data augmentation can make the model more robust in
76 heterogeneous and low-resource situations.
- 77 • To evaluate the performance of ConvNeXt compared to other state-of-the-art
78 architectures such as ResNet and Swin Transformer in terms of accuracy and
79 computational efficiency.

80

81 In Figure 1, it is observed that the proposed ConvNeXt, which is a modernised convolutional
82 neural networks architecture, outperforms other current complex models such as ResNet, DeiT,
83 and Swin Transformer in both accurate training and computationally efficient when trained on
84 ImageNet datasets (8). ConvNeXt applies to many fields due to its scalability and enhanced
85 performance obtained through training on extensive datasets like ImageNet-22 (9).

86

87 Figure 1: Performance of ConvNext on ImageNet. Source: (8)

88

89 Given the scarcity of big medical-labelled datasets, transfer learning and data augmentation
90 have been extensively used in this work. Transfer learning is a technique that enables the
91 models to use the knowledge learned from one task and apply it to another task, such as malaria
92 detection from images. Conversely, the data augmentation technique enhances the dataset's

93 size by manipulating the image conditions to make the model less sensitive to practical
94 variations.

95

96 This study explores a novel application of the ConvNeXt architecture, combining transfer
97 learning with advanced data augmentation techniques for medical imaging. In contrast to
98 conventional deep learning, feature extraction capabilities can be improved by transferring
99 knowledge from the large-scale ImageNet dataset despite the relatively small dataset.
100 Moreover, integrating explainable AI tools like LIME and LLaMA brings a unique dimension
101 to a diagnostic process where the model's decision-making can be visually and textually
102 interpreted. These innovations highlight the opportunities for AI-driven systems to enhance
103 diagnosis accuracy while promoting greater clinician acceptance by being transparent in AI
104 predictions.

105

106 **2. Materials and Methods**

107

108 2.1 Dataset Acquisition

109

110 The dataset used for training and evaluating the deep learning models in this study was obtained
111 from the Lister Hill National Center for Biomedical Communications (LHNCBC), part of the
112 National Library of Medicine (NLM), which hosts a publicly available collection of malaria-
113 infected blood smear images, available at (10). This dataset was initially collected at the
114 Chittagong Medical College Hospital in Bangladesh, and the data are made up of thin blood
115 smear images (11). The dataset comprises 27,558 images evenly distributed between
116 parasitised and uninfected samples, with each category containing 13,779 images.

117

118 This balanced distribution ensures the reliability and validity of subsequent analyses and
119 research findings (12). It prevents the model from being trained to overemphasise one class
120 over the other in cases where the dataset could be more balanced between the two categories.

121

122 The blood smear images were photographed using a smartphone camera, which was held up to
123 the eyepiece of a microscope; this configuration mimics the conditions likely to be found in
124 low-resource settings. All the images were obtained from the blood smears stained with giemsa,
125 the standard method of malaria diagnosis by microscopy. The images were then reviewed and
126 labelled by expert technicians to determine whether or not the parasites were present. Each
127 image had a dimension of 5312x2988 pixels, with the circular area depicted as the view through
128 the microscope lens. As a result of the limited resources used in the imaging of blood smears
129 and the variability in the preparation of the samples, the images had variations in lighting,
130 contrast, and colour balance.

131

132 To avoid any possibility of identifying patients who may be reflected in the dataset, the patient
133 data were de-identified before online publication. The Institutional Review Board (IRB)
134 approved using the data at the NLM (IRB#12972)(11). Additionally, this study received ethical
135 clearance from the University of Johannesburg's School of Consumer Intelligence and
136 Information Systems Research Ethics Committee (SCiSREC) under ethical clearance code
137 2024SCiS040. This clearance is valid for three years, starting on 1 August 2024. This approval
138 ensures that the use of the data complies with the set ethical procedures for handling and
139 analysing medical data.

140

141 In Table 1, the image counts are supplemented with both categories' mean and standard
142 deviation.

143 Table 1: Image Statistics for Parasitised and Uninfected Categories

Category	Image Count	Mean (R, G, B)	Standard Deviation (R, G, B)
Parasitised	13,779	[0.4507, 0.3882, 0.3955]	[0.3109, 0.2954, 0.2656]
Uninfected	13,779	[0.4478, 0.4841, 0.4634]	[0.2966, 0.3222, 0.2919]

144

145 The mean and standard deviation values are pixel intensities, which are the extent of brightness
146 or colour of a pixel on an image (13). The pixel intensity values are from red, green, and blue
147 (RGB). These channels amount to the image's colour, which ranges from 0, representing black,
148 to 1, representing white. These are colour measurements for the pictures and consist of average
149 colour intensities of the parasitised and uninfected samples regarding brightness and colour
150 changes (14). As the value in a channel increases, the corresponding pixel intensity value will
151 increase, implying a light and or intense colouration.

152

153 Pixel intensity analysis has been applied to detect parasites within microscopic images,
154 particularly in identifying infected and uninfected cells. Studies such as (14) have shown the
155 use of pixel intensities for classifying stained microscopy images based on colour variations.
156 Similarly, several studies have demonstrated that pixel intensities are useful in diagnosing
157 parasitic conditions (15-18). These works provide solid grounds for using pixel intensities,
158 particularly in automated diagnostic systems.

159

160 The average pixel intensity of the colour channels in the "Parasitised" images is Red = 0.4507,
161 Green = 0.3882, Blue = 0.3955. This indicates that the 'Parasitised' images have moderate
162 pixel intensity levels in these channels. The standard deviations for the same set of images are
163 approximately 0.3109, 0.2954 and 0.2656, indicating the pixels' spread of intensity values. A
164 larger standard deviation indicates greater dispersion; in this case, the dispersion is relatively

165 moderate across the colour channels, which means there is some difference in image quality or
166 staining intensity but not too extensive.

167

168 The pixel intensities for the “Uninfected” images are 0.4478, 0.4841, and 0.4634 in the three
169 colour channels. These values show slightly higher pixel intensity, especially in the Green and
170 Blue channels, than the “Parasitised” images. The standard deviations for the “Uninfected”
171 images, which are 0.2966, 0.3222, and 0.2919, also indicate that the pixel intensity is
172 moderately dispersed, similar to the “Parasitised” images but with variability across the
173 channels. This dataset is well-balanced and statistically stable, which allows for the practical
174 training of malaria detection models.

175

176 **2.2 Image Pre-processing**

177

178 2.2.1 Image Resizing

179

180 The data used in this study were pre-processed before being applied to the deep-learning
181 models used in this study. The input size of the original images at a resolution of 5312x2988
182 pixels offered an abundance of visual information; however, they were computationally costly
183 and were not appropriate for the input dimensions of most current deep learning architectures.
184 To this end, all images were resized to 224×224 pixels as most of the current ConvNeXt, Swin
185 Transformer, and ResNet models have a standard input image size of 224×224 pixels (19).

186

187 These steps are essential, especially in resizing the images, because this has an impact on the
188 speed with which the models can process the data; it lowers the memory usage during the
189 process and allows the training of the models to be more efficient without losing much detailed

190 information that is important in identifying malaria parasites. However, in most cases, when
191 images are resized, they tend to lose some details, which can be crucial, especially when the
192 resizing is performed to a great extent, such as here (20).

193

194 To avoid loss of information during image resizing, the Lanczos interpolation filter was used
195 during the resizing of the image in Python, a method used for image scaling when the quality
196 of the image should be preserved when downscaled (21, 22).

197

198 Lanczos interpolation is employed to resize the images while preserving the edges and fine
199 details of the images by using the ‘sinc’ function so that the morphology of malaria parasites
200 is well retained in the resized images (21, 23). Thus, applying this method, the details of the
201 pre-processed images were preserved, which is crucial for accurately detecting and classifying
202 the parasitised and uninfected blood smears by the deep learning models. Hence, despite the
203 lower image resolution, which helped to enhance computational speed, the Lanczos filter was
204 used to maintain an adequate image quality for the intended application.

205

206 Sinc or sinus cardinalis is a mathematical function applied in signal processing, especially in
207 image processing for interpolation, especially in resampling or resizing images (24). The sinc
208 function is defined as:

209

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \quad (1)$$

210

211

212 The sinc function, shown in Equation 1, is a continuous and periodic function defined by
213 oscillations and decreases when moving away from zero (25). The sinc function is essential in

214 Fourier analysis and can be described as the ideal low-pass filter in the frequency domain. It
215 avoids the problem of aliasing, where different signals appear identical when sampled and
216 produce distorted resampled images or signals.

217

218 This study uses the sinc function in Lanczos interpolation, which calculates new pixel colours
219 while an image is being scaled. The sinc function decays and can be used as a weighting
220 function when resampling surrounding pixel values, which produces much smoother
221 transitions between pixels (26). This allows high-frequency details, such as sharp edges, to be
222 kept when the image size is decreased, which is essential when dealing with the images of
223 blood smears used in malaria detection.

224

225 2.2.2 Image normalisation

226

227 After resizing the images, the normalisation process was performed to ensure that pixel
228 intensities were consistent in the dataset. Normalisation is an essential procedure in deep
229 learning since it enhances the stability of the learning process by preventing variations in the
230 properties of the input data (27, 28). For the normalisation, the mean and the standard deviation
231 of the pixel values of the images in each of the three colour channels, namely the red, the green
232 and the blue, were computed.

233

234 The numerical results of the calculated mean values are listed in Table 1, are 0.4507 for the red
235 channel, 0.3882 for the green channel and 0.3955 for the blue channel; the standard deviation
236 values are 0.3109, 0.2954 and 0.2656, respectively. These values were then applied to
237 normalise each image to ensure their pixel values had a mean around zero and the variance was
238 one. This pre-processing step is essential to aid the models in learning and performing well

239 when faced with data they have not encountered before by normalising the brightness and
240 contrast of the images.

241

242 2.3 Data augmentation

243

244 Data augmentation is significant because it improves the transferability of the deep-learning
245 models for malaria parasite identification (29). The lack of large, high-quality datasets is a
246 common challenge in low-resource settings. Thus, data augmentation is useful in artificially
247 increasing the amount of available data and introducing variability typical in real-life
248 conditions (30, 31). To definite the models and make them capable of identifying the malaria
249 parasites in varied situations, many data augmentation operations were performed on the base
250 images, thereby creating a large dataset.

251

252 In this study, the techniques used for data augmentation were used for imitating diverse
253 conditions under which blood smear images may be taken. Horizontal flipping applied the
254 transformation that reflected the images across the x-axis, and vertical flipping was the
255 transformation that reflected the images along the y-axis (32). These steps were critical to
256 enable the model to learn how to identify malaria parasites, irrespective of their orientation. In
257 the case of blood smear microscopy, the placement of the smears is only sometimes consistent.

258

259 Rotation is another transformation, where images were rotated at an angle of 45 degrees (33,
260 34). While preparing and analysing blood smears, they may not always be appropriately placed;
261 hence, the model should be able to deal with rotated images. Scaling was used to control the
262 size of the images with a scaling factor of 0.5 to 1.5 (33). This enhancement was functional in
263 modelling the different magnification levels because the observed visual characteristics may

264 range in size based on the microscope's settings. This is because by training on scaled images,
265 the model is in a better position to detect parasites regardless of their size.

266

267 Gaussian noise was added to the images to simulate potential noise observable in real-world
268 image acquisition. The noise levels varied from 0.01 to 0.05 times the maximum pixel intensity
269 value, which is 255 (33, 35). This made the training images similar to those taken in conditions
270 that are not favourable, such as low light or inadequate focus of the camera. Contrast
271 adjustments were made, whereby the contrast values were changed by a factor of 0.8 to 1.2
272 (33). Using contrast-enhanced images, the model could learn and distinguish parasites even in
273 different image contrasts, thus improving its performance in different diagnostic conditions.

274

275 Besides these fundamental transformations, some more specific affine transformations were
276 used, such as shearing (33), which changes the image along the x- or the y-axis to mimic some
277 shifting or distortion that may occur while preparing the slides. Several techniques of blurring
278 the images were used to make the images appear out of focus. Gaussian blurring with a sigma
279 range of 0.0 to 3.0 was applied to the images (36). These techniques imitated the conditions of
280 blurred images due to improper focus on the microscope, and thus, the model learned to
281 recognise parasites in somewhat blurred images.

282

283 Sharpening was applied to the images to make the details more apparent, with the alpha range
284 from 0 to 1 and lightness range from 0.75 to 1.5 (37). Sharpening improves the details in the
285 images, especially the edges of the parasites, which are very important in identification.
286 Colour-based augmentations, such as changes to hue and saturation, were also incorporated
287 (38). These adjustments were informed by differences in staining method when preparing blood
288 smears and differences in lighting or imaging systems. Elastic transformation and dropout were

289 also applied, introducing more variability. The elastic deformations, which were controlled by
290 $\alpha=50$ and $\sigma=5$, caused minor shifting of the image, and this was beneficial as the model
291 could identify parasites even if there was a slight deformation in the image (39, 40).

292 Cropping was performed randomly by removing parts of the image with 0.02 to 0.1 pixels (41).
293 This transformation aided in the model being more generalised so that it could still predict the
294 images even if some parts of the images were cut or blurred. Channel shuffling was applied
295 with a probability of 0.35, where the colour channels of the images were randomly
296 interchanged to cope with colour channel variations (42).

297

298 Applying these transformations increased the dataset to 606,276 samples, where 303,138 is the
299 number of parasitised samples and 303,138 is the number of uninfected samples. The mean
300 pixel values for the parasitised images were proposed to be [0.44839746, 0.38788548,
301 0.39583275] while those of the uninfected images were [0.4479274, 0.4817927, 0.46273437].
302 The standard deviations were 0.30283427, 0.28804937 and 0.26085448 for the parasitised
303 category, while those of the uninfected category were 0.2897945, 0.313396 and 0.285556.
304 These statistics demonstrate the effects of the augmentation process that made the dataset more
305 extensive and diverse in terms of the images' features.

306

307 Figure 2 visually represents the dataset after data augmentation, explicitly comparing the
308 number of images in two categories: The two main groups used in this study were parasitised
309 and uninfected. There are two categories in total, each of them including about 303,138 images,
310 proving the balanced distribution of the dataset after the augmentation process. This is very
311 important in training the machine learning models, particularly in classification problems, to
312 avoid being influenced by one category due to data sampling.

313

314 Figure 2: Image count per category (After data augmentation)

315

316 The augmentation made it possible to introduce variability in the form of geometric
317 transformations, noise, scaling and colour variations, among others, as shown in Figure 3. This
318 was performed by adding more training images and mimicking many scenarios one might
319 encounter when capturing blood smears in practice, thus improving the generalisation
320 capability of the models. Such augmentation aims to ensure that the models can identify the
321 malaria parasites identifiable across various images and illumination and rotations, which are
322 common in real-world applications.

323

324 Figure 3: Sample images after data augmentation

325

326 Table 2 summarises the basic descriptive statistics of the characteristics of the augmented
327 dataset used to train the malaria detection models. As seen in Table 2, the statistics of the
328 dataset are reasonably even, and pixel intensity and variation are slightly different between
329 Parasitised and Uninfected images, which will help the model during training. These colour
330 distribution patterns are significant as they enable the model to distinguish between the
331 parasitised and uninfected blood smear images, thus enhancing the model's performance.

332

333 Table 2: Image Statistics for Parasitised and Uninfected Categories (After Augmentation)

Category	Image Count	Mean (R, G, B)	Standard Deviation (R, G, B)
Parasitised	303,138	[0.44839746, 0.38788548, 0.39583275]	[0.30283427, 0.28804937, 0.26085448]
Uninfected	303,138	[0.4479274, 0.4817927, 0.46273437]	[0.2897945, 0.313396, 0.285556]

334

335

336 2.4. Algorithms

337

338 This paper utilised several sophisticated deep-learning models to identify malaria parasites in
339 microscopic blood smear images. These models cover various architectures that effectively
340 capture local details, e.g., parasite shapes, and global information, e.g., cell distributions, in the
341 images. The models used were the Swin Transformer (Swin Tiny), ResNet18, ResNet50,
342 ConvNeXt Tiny, ConvNeXt V2 Tiny, and a modified version of ConvNeXt V2 called
343 ConvNeXt V2 Remod. Every architecture has been chosen to work with high-resolution
344 medical images and, at the same time, employ transfer learning, allowing the model to take
345 knowledge from previously trained models trained on large datasets like ImageNet.

346

347 Swin Transformer Architecture

348

349 This study used the Swin Transformer, particularly the Swin Tiny model, to exploit its window-
350 based multi-head attention (43). This architecture splits the input images into multiple non-
351 overlapping local windows and then learns the self-attention from these local windows. This
352 way, Swin Transformers can capture regional and global information in images, which is
353 beneficial for tasks such as medical image analysis (44). The Swin Transformer architecture,
354 as shown in Figure 4, is where images are divided into a sequence of non-overlapping patches.
355 Then, a linear embedding layer is applied to generate patch tokens. This architecture is divided
356 into four stages. Each stage consists of several Swin Transformer blocks and a patch merging
357 layer, which reduces spatial dimensions while increasing the feature dimensions, allowing for
358 both local and global context to be effectively captured across the entire image.

359

360 Figure 4: Architecture of a Swin Transformer. Source (43)

361 Mathematically, the self-attention mechanism within a Swin Transformer can be represented
362 as:

363

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

364

365 where Equation 2 represents the queries, keys, and values, respectively, and d_k is the dimension
366 of the keys. The model incorporates a relative position bias term that enhances its ability to
367 encode the spatial structure within each window.

368

369 Swin Tiny was pre-trained on ImageNet-1K, a dataset containing over a million labelled
370 images. It achieved a top-1 accuracy of 81.2% and a top-5 accuracy of 95.5% on ImageNet,
371 with 28 million parameters and a computation cost of 4.5 GFLOPs (45).

372

373 ResNet Architectures

374

375 The ResNet18 and ResNet50 were selected as the baseline models to be compared with the
376 proposed models. These models are a part of the Residual Networks, which are the types of
377 neural networks developed to address the vanishing gradient problem that is a big challenge in
378 training deep neural networks (46). The residual block is central to ResNet's architecture and
379 is expressed mathematically as:

380

$$y = F(x,(W_i)) + x \quad (3)$$

381

382 where x is the input, $F(x, (W_i))$ is the learned transformation with weights W_i and y is the block
383 output, as shown in Equation 3. This structure helps the network to learn identity mapping
384 when deeper layers do not help enhance performance and allow a flow of gradients across many
385 layers.

386

387 ResNet18 is a less complex model with 18 layers, while ResNet50 is a deeper model with 50
388 layers; hence, it can learn more complex data features. ResNet18 was pre-trained on ImageNet
389 and fine-tuned for malaria detection, and the second model was pre-trained on ImageNet and
390 fine-tuned for malaria detection (47). Although ResNet18 was used as the baseline model,
391 ResNet50 had a deeper network and could thus identify more intricate visual patterns in the
392 blood smear images and differences between them (48).

393

394 ConvNeXt and ConvNeXt V2 Architectures

395

396 ConvNeXt Tiny, a novel architecture based on conventional convolutional neural networks
397 (CNNs), was another critical architecture used in this study (49, 50). Also based on Vision
398 Transformers (ViTs), ConvNeXt is a model that combines the hierarchical architecture to
399 incorporate both high and low-level features of images. Consequently, the feature extraction of
400 Swin Transformer, ResNet, and ConvNeXt model block designs are different, as shown in
401 Figure 5. Swin Transformer block captures local and global features with multi-head self
402 attention (MSA) with shifted windows ($w7 \times 7$), followed by layer normalisation and Gaussian
403 Error Linear Unit (GELU) activation. A ResNet block is designed based on the residual
404 structure of 1×1 and 3×3 convolutions, batch normalisation and Rectified Linear Unit (ReLU)
405 activation to help feature learning. ResNet is modernised to ConvNeXt block by replacing 3×3

406 convolutions with depthwise convolutions (d7x7) with an extra parameter of channel
407 multiplier, as well as using layer normalisation and GELU activation for better efficiency.

408

409 Figure 5: Block designs for the used models. Source (8).

410 The core operation in ConvNeXt is convolution, mathematically described as:

$$(S * K)(i,j) = \sum_m \sum_n S(i + m, j + n)K(m,n) \quad (4)$$

411

412 In Equation 4, S represents the input matrix (image), K, in Equation 4, is the convolutional
413 kernel, and (i,j) denotes spatial positions in the image. ConvNeXt Tiny model was pre-trained
414 on ImageNet, with the top-1 accuracy of 82.1% and has 28M parameters and 4.5 GFLOPs of
415 computation. The pre-trained model was downloaded from GitHub and then fine-tuned to the
416 malaria dataset to teach the model the features of detecting malaria parasites.

417

418 Following the success of ConvNeXt, ConvNeXt V2 had additional architectural enhancements,
419 including learning rate scheduling and modified activation functions for improving the image
420 classification task performance (50). The ConvNeXt V2 Tiny model employed in this study,
421 which was also pre-trained on ImageNet, provided a top-1 accuracy of 83.0 per cent and had a
422 computational complexity of 4.47 GFLOPs with 28.6 million parameters. Similar to
423 ConvNeXt V2, the model was fine-tuned on the malaria dataset to change the pre-trained
424 weights of the model to adapt to the characteristics of the images of parasitised and uninfected
425 blood smears.

426

427 ConvNeXt V2 Remod

428

429 This study's ConvNeXt V2 Remod model was based on the ConvNeXt V2 Tiny architecture.
430 In the training process, label smoothing was used with $\alpha=0.1$ (51, 52). Label smoothing shifts
431 the target labels to prevent the model from attaining extreme confidence and helps deter the
432 overfitting of the model (52). Mathematically, label smoothing modifies the loss function as
433 follows:

434

$$L_{CE} = - \sum_{i=1}^N \left((1 - \alpha)y_i + \frac{\alpha}{C} \right) \log(\hat{y}_i) \quad (5)$$

435

436 where C is the number of classes (in this case, $C=2$ for binary classification), α distributes
437 a small part of the probability mass to the incorrect classes, as shown in Equation 5.

438 The block structures of ConvNeXt V1 and V2, shown in Figure 6, show the core improvements
439 made from one version to the other. The ConvNeXt V1 block consists of depthwise
440 convolutions ($d7 \times 7$) and then layer normalisation (LN) and GELU activation for efficient
441 feature extraction. The first version of the model introduced LayerScale, which helped stabilise
442 the process. On the other hand, the ConvNeXt V2 block keeps the heart of V1 but introduces
443 Generalised ReLU Normalisation (GRN). This new state-of-the-art normalisation technique
444 enhances the model's stability and efficacy. This addition of GRN with the elimination of
445 LayerScale boosts the generalisation ability of ConvNeXt V2 across several tasks.

446

447 Figure 6: ConvNeXt block designs. Source (50).

448

449 The AdamW optimiser, shown in Equation 6, was selected for this task as it is a version of the
450 Adam optimiser with weight decay for better generalisation and reduced overfitting (53, 54).
451 AdamW updates the model's weights using first- and second-order moments of the gradients.

452 The main distinction between Adam and AdamW is that in AdamW, the weight decay is not
453 applied directly to the gradients, which allows the magnitude of the weights to be preserved.

454

455 The learning rate was set to 0.0005, and a weight decay of 0.01 was applied to penalise large
456 weights, thus avoiding overfitting the training data (55). The AdamW optimiser is defined as:

457

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (6)$$

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

458 where:

459 g_t is the gradient of the loss with respect to the parameters at step t

460 m_t and v_t are the moving averages of the gradient and its square, respectively,

461 θ_t represents the model parameters at step t

462 α is the learning rate, and

463 β_1 , β_2 and ϵ are hyperparameters of the optimiser.

464

465 To improve learning efficiency, the OneCycleLR scheduler was applied (56). This scheduler
466 modifies the learning rate throughout training to balance exploration (high learning rate) and
467 refinement (low learning rate). The learning rate is first set to maximum value and then

468 decreases over time, which helps the model avoid local minimum in the initial training phase
469 while fine-tuning the weights when training continues.

470

471 This is set to zero initially, rises to a maximum of 0.0005 and then decreases using a cosine
472 annealing schedule, shown in Equation 8, over ten epochs. The cosine annealing formula is
473 given as Equation 7 by:

474

$$\eta_t = \eta_{min} + 0.5(\eta_{max} - \eta_{min}) \left(1 + \cos \left(\frac{t}{T} \pi \right) \right) \quad (7)$$

475 where:

476 η_t is the learning rate at time step t ,

477 η_{max} and η_{min} are the minimum and maximum learning rates, and

478 T is the schedule's total number of time steps (epochs).

479

480 To optimise the training of the NVIDIA[®] Tesla[®] P100 graphics processing unit (GPU) used in
481 Kaggle, PyTorch's automatic mixed precision (AMP) was implemented to train in mixed
482 precision. It is a technique where half of the computations are performed in a 16-bit floating-
483 point format, whereas the other half are performed in a 32-bit floating-point format. The
484 gradients are calculated in the FP16 format for the sake of optimisation, while the master
485 weights are stored in the FP32 format for the sake of numerical accuracy.

486

487 For this, gradient scaling was used to deal with the differences between FP16 and FP32.

488 Gradient scaling is the process of scaling the gradients by a specific factor to prevent the
489 gradients from being too small (and thus causing underflow) or too large (which will cause

490 overflow). This scaling assists in controlling the training process, especially for deep networks
491 with numerous parameters to be estimated.

492

493 The training of the ConvNeXt V2 Remod model was performed for ten epochs, with the
494 learning rate being adjusted in a way that aims at deriving the highest performance. The model
495 was trained with high performance and reliability using mixed precision through automatic
496 mixed precision and with the help of checkpointing. Thus, the optimisation strategies and the
497 advanced architectural features helped to enhance the recognition of malaria parasites in blood
498 smear images at the end of the training process.

499

500 2.5 Model Development

501

502 2.5.1 Transfer learning

503

504 This work applied transfer learning, where pre-trained models were obtained from GitHub
505 repositories. Transfer learning offered several advantages as it significantly cut down training
506 time. As the models had previously been trained on large datasets such as ImageNet, the
507 computational power and time needed for fine-tuning the models for malaria classification
508 were considerably lower than that required for training new models. Secondly, transfer learning
509 played a role in reducing overfitting, which is a big problem when dealing with limited samples.
510 Through pre-training, the models with a large set of images, continued to perform well on the
511 malaria classification task due to their generalisation ability.

512

513 In addition, transfer learning was used to ensure the models' accuracy and robustness were not
514 compromised by the new task of distinguishing between parasitised and uninfected blood

515 smears since the models could still rely on their prior experience with identifying general
516 objects in images (57). This approach was instrumental in resource-limited settings where
517 access to vast amounts of labelled data and powerful computing resources is limited.

518

519 Swin Transformers, ResNet 18, ResNet 50, ConvNeXt Tiny and ConvNeXt V2 Tiny were the
520 models employed in this study, all pre-trained on large datasets of generic images. There are
521 several benefits of using these pre-trained models; first, they are already equipped with feature
522 extraction capabilities that can be fine-tuned for a specific task, for instance, malaria parasite
523 detection in microscopic blood smear images (58-60).

524

525 The pre-trained Swin Transformer models were acquired from the repository managed by
526 Microsoft and are accessible at (61). Swin Transformers are well-suited for handling high-
527 resolution images due to their unique mechanism of dividing images into non-overlapping
528 windows, enabling efficient computation of self-attention within each window (43, 62). This
529 approach makes it possible for the model to learn both the image's local and global features,
530 which is very important during the classification, especially in tasks such as parasite
531 identification.

532

533 ConvNeXt and ConvNeXt V2 models used in this work were obtained from the public
534 repositories established by Facebook AI Research. ConvNeXt models can be accessed at (63).
535 They represent advancements over regular convolutional neural networks (CNNs) with
536 features derived from Vision Transformer architectures (8, 49). ConvNeXt V2, accessible at
537 (64), incorporates improvements that include better resource management and improved
538 feature extraction and, hence, is even more suitable for tasks such as malaria classification (50).

539

540 These models were first pre-trained on ImageNet, one of the most commonly used datasets in
541 the computer vision field (65). The ImageNet dataset is one of the most extensive, with over
542 14 million labelled images spanning 1,000 categories of objects, including animals, buildings,
543 nature, and food, amongst others (66). This large amount of data helps these models to learn
544 the features well. Hence, they are ideal for transfer learning applications, such as detecting
545 malaria parasites with limited datasets.

546

547 Fine-tuning these models for classifying parasitised and uninfected blood smears involved
548 adjusting the final fully connected layers to output two classifications: The blood smears were
549 either parasitised or uninfected (67). It is essential to replace the pre-trained models' last layers
550 with new layers specifically designed for malaria parasite detection. While the pre-trained
551 models are well-trained in the general characteristics of the images, such as edges, patterns,
552 and textures, the final layers can be focused on learning the unique characteristics of the blood
553 smears relevant to malaria diagnosis. In this case, the final layers were replaced with custom
554 layers designed for binary classification, parasitised versus uninfected.

555

556 2.5.2 Training Procedure

557

558 The training process for this study was performed on Kaggle, which offered the use of
559 NVIDIA® Tesla® P100 GPU. This is mainly because of the GPU's architecture and
560 computational ability for deep learning. The Tesla® P100 is based on NVIDIA®'s Pascal
561 architecture, includes 16GB of HBM2 memory and performs 10.6 TeraFLOPS in single-
562 precision floating-point operations per second. Such a computational capability helped handle
563 large image datasets used in this study without experiencing low speeds. The P100 had enough

564 memory to operate on large and high-resolution images so that batch processing of models
565 could be performed faster during training.

566

567 The training was performed using mixed-precision training with the help of PyTorch's
568 automatic mixed precision (AMP) (68). AMP is a technique which enables deep learning
569 models to train faster and with less memory than standard training (68, 69). AMP works on
570 some parts of the network. For instance, the convolutions are computed using 16-bit floats,
571 while the rest of the model's weights and gradients are calculated using 32-bit floats to balance
572 between speed and accuracy.

573

574 This technique also enhances the speed of training and optimisation of the GPU, especially the
575 Tesla[®] P100, which performs operations on large datasets with great accuracy. Furthermore,
576 the vanishing and exploding gradient issues that are evident when training deep networks with
577 vast parameter spaces like ConvNeXt or Swin Transformer models are also solved by AMP.

578

579 Optimisation Strategy

580

581 This study opted to use the Adam optimiser, as shown in Table 3, which has been known to be
582 efficient in adjusting the learning rate during training (70). Adam uses the benefits of AdaGrad
583 and RMSProp optimisations involving gradient averages and second-order moments to adjust
584 the model weights (53, 71). This is advantageous for Adam for handling the sparse gradients
585 and noisy data, which are familiar with medical image data. The learning rate of the optimiser
586 was set to 0.0001 for most models, including ConvNeXt Tiny, ResNet18, and ResNet50, due
587 to its ability to fine-tune deep learning features, especially for medical imaging data that
588 typically contain noise and sparse gradients (72). A low learning rate is essential to guarantee

589 that the model gives only small weight adjustments to make the best estimate without
 590 overfitting highly sensitive models on the loss surface (73). For the ConvNeXt V2 Remod
 591 model, a higher learning rate of 0.0005 was used alongside weight decay of 0.01 to improve
 592 model generalisation (74).

593

594 Table 3: Hyperparameters used for fine-tuning different models

Model	Learning Rate	Batch Size	Optimiser	Weight Decay	Loss Function	Scheduler	Epochs	Mixed Precision
ConvNeXt Tiny	0.0001	128	AdamW	Not Used	Cross Entropy Loss	None	10	GradScaler & autocast
Swin Transformer	0.0001	128	Adam	Not Used	Cross Entropy Loss	StepLR	10	GradScaler & autocast
ResNet18	0.0001	32	Adam	Not Used	Cross Entropy Loss	None	10	GradScaler & autocast
ResNet50	0.0001	128	Adam	Not Used	Cross Entropy Loss	None	10	GradScaler & autocast
ConvNeXt V2 Tiny	0.0001	128	AdamW	Not Used	Cross Entropy Loss	None	10	GradScaler & autocast
ConvNeXt V2 Remod	0.0005	128	AdamW	0.01	Cross Entropy Loss (Label Smoothing = 0.1)	OneCycleLR (Cosine)	10	GradScaler & autocast

595

596 The adaptive learning rate of Adam makes the model converge faster than the normal stochastic
 597 gradient descent (SGD) (75). In this study, it was beneficial for large and elaborate models

598 such as ConvNeXt and Swin Transformer with many parameters. Adam allowed the model to
599 automatically and adaptively set the learning rate for each parameter, which made the model
600 optimise itself and increase the training process's convergence rate.

601

602 Gradient Scaling and Stability

603

604 Gradient scaling was used during the optimisation to avoid training process instabilities (76).

605 It is a method applied in mixed-precision training to prevent the gradients from becoming too

606 small (vanishing gradients) or too large (exploding gradients) (77). In large networks, such as

607 those used in this study, the parameter space is also significant, which can result in oscillations

608 during training. Applying gradient scaling before backpropagation helped avoid the potential

609 numerical precision problems common in deep learning models.

610

611 A learning rate scheduler was used to stabilise the training process during implementation. A

612 learning rate scheduler adjusts the learning rate during training according to the given

613 performance of a model (78, 79). In this case, the scheduler had a learning rate reduction of 0.1

614 every seven epochs if the validation loss did not decrease. This technique is very helpful in

615 preventing the model from becoming trapped in the local minima. When the validation loss

616 stalls, reducing the learning rate helps the optimiser fine-tune the model weights and allows

617 the model to perform better and converge. For ConvNeXt V2 Remod, a OneCycleLR (cosine)

618 scheduler was applied to balance fast convergence with the risk of overfitting

619

620 Loss Function: Cross-Entropy Loss

621

622 The loss function used for this study was the Cross-Entropy Loss function, which is most
623 suitable for the binary classification of the malaria parasite cases (parasitised and uninfected
624 red blood cells) (80). Cross-entropy loss measures how far the predicted class probabilities are
625 from the actual class labels (81). It is more useful in classification as it penalises the predicted
626 output that is not in line with the actual output. In the case of ConvNeXt V2 Remod, label
627 smoothing was also applied to the cross-entropy loss to mitigate the effect of noisy labels and
628 improve generalisation. In this work, the model was trained to predict the likelihood of a given
629 blood smear image being parasitaemic. The output from the final layer in the model is a vector
630 of predicted probabilities, which are compared to the actual class labels, and the weights of the
631 model are updated to minimise this loss function known as Cross-Entropy Loss.

632

633 Mathematically, Cross-Entropy Loss is computed as:

634

$$Loss = - \sum_{i=1}^N [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \quad (8)$$

635

636 where:

637 N is the number of samples,

638 y_i is the true label (either 0 for uninfected or 1 for parasitised),

639 p_i is the predicted probability for the corresponding class.

640

641 In this context, Cross-Entropy Loss, displayed in Equation 8, helps the model to give a high
642 probability to the correct class and a low probability to the incorrect class (82). Since the dataset
643 has equal parasitised and uninfected images, Cross-Entropy Loss prevents the model from
644 leaning towards one class and produces equally good results.

645

646 2.6 Model Application and Deployment

647

648 To enhance the applicability of the trained ConvNeXt models, a web application was developed
649 using Gradio to facilitate real-time identification of malaria parasites in blood smear images.
650 The Python library Gradio designed application has an interface where users can upload blood
651 smear images and receive real-time diagnoses in a limited resource environment. It is
652 accessible at (83).

653

654 The application (app) was designed with two main functionalities: The proposed solution
655 includes image classification using ensemble models and the explanation of the predictions
656 using the LIME model (Local Interpretable Model-agnostic Explanations) and Llama 3.1 by
657 Facebook Research (84-87).

658

659 The application presents novel contributions through the use of an ensemble of deep learning
660 models (ConvNeXt Tiny and ConvNeXt V2) that help improve diagnosis performance and
661 reliability. To do this, the app takes the average of these models' predictions by assigning
662 weights to each of the models to make a more accurate determination on whether a blood smear
663 is parasited or not. This ensemble approach outperforms the conventional single-model systems
664 employed in medical diagnostics to provide a more accurate result, especially in the limited
665 resource environment (88). Furthermore, the app employs mixed-precision training with the
666 help of GradScaler, which ensures the high performance with the minimal consumption of
667 resources, which will be beneficial for the areas with poor infrastructure.

668

669 One of the main aspects of the app is the explainability element, using LIME for visualisation
670 and LLaMA for textual elaboration (87, 89). LIME gives out visual maps that demonstrate the

671 parts of the image which are used in making the decision, which helps in improving the
672 explainability of the predictions to medical practitioners (89). LLaMA enables an additional
673 step in explaining the results as it generates human-readable descriptions of the context and,
674 therefore, helps interpret the machine's predictions (87). These explainability features together
675 with efficiency in the use of resources make the Malaria Diagnosis App not only unique but
676 also very useful for use in malaria endemic areas.

677

678 2.6.1 Use of best performing models

679

680 The application combines the strengths of two fine-tuned models, namely ConvNeXt Tiny and
681 ConvNeXt V2 Tiny Remod, through the ensemble method (90). This approach combines the
682 prediction of the first model with the second model's prediction to make the final decision,
683 increasing the overall reliability of the decision made (91). From the two models used in this
684 study, ConvNeXt Tiny was purposely selected for its computation efficiency, which allows it
685 to process images at high speed. At the same time, ConvNeXt V2 Tiny Remod was chosen
686 because of its higher accuracy and precision, as highlighted in the comparative analysis.

687

688 Every image the user uploads is first processed through a pre-processing step, which includes
689 resizing the image to 224 x 224 pixels and normalising the image using the mean and standard
690 deviation of the malaria dataset. This check helps meet the conditions the trained models expect
691 as input data. The image is then passed through both models, and the average output of the two
692 models is computed. The final decision on whether a blood smear is parasitised or uninfected
693 is based on the average of the two models presented in this work. This approach enhances the
694 diagnostic performance and decreases the rates of false-positive results, thus making the
695 diagnosis more accurate.

696

697 2.6.2 Model Explanation with LIME

698

699 To further enhance the transparency and interpretability of the malaria parasite detection
700 models, the LIME (Local Interpretable Model-agnostic Explanations) algorithm was
701 employed. LIME works by perturbing the input data - in this case, the blood smear images -
702 and observing how the model's predictions change in response to these perturbations. The
703 algorithm generates explanations by locally approximating the model's decision boundary and
704 identifying the regions within the image that most influence the final classification. This
705 process is precious for understanding deep learning models, often called "black boxes."

706

707 Mathematically, LIME can be understood as follows:

708

709 Using a kernel function, LIME approximates the complex model around the local region of the
710 explained instance: $\pi(x_0, x')$.

711

712 LIME minimises the following loss function to generate explanations:

713

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (9)$$

714 where:

715 $\xi(x)$ is the explanation for the instance x ,

716 G is the class of interpretable models, $\mathcal{L}(f, g, \pi_x)$, in Equation 9, is the loss function that

717 measures the fidelity of the interpretable model g to the complex model f in the local region

718 defined by the kernel π_x .

719 Omega(g) is a complexity term that penalises the complexity of the interpretable model g,
720 ensuring that the explanation remains simple and human-readable.

721

722 The kernel function $\Pi(x_0, x')$ assigns higher weights to perturbed samples closer to the original
723 instance x_0 ensuring that the explanation focuses on the local behaviour of the model. In the
724 case of image classification, this kernel function is often defined based on the Euclidean
725 distance between the perturbed image x' and the original image x_0 .

726

727 Once the simpler model $g(x)$ is trained to highlight the regions in the image most responsible
728 for the model's prediction. These highlighted regions correspond to the image features—such
729 as ring-shaped structures or abnormal cell morphologies that the model associates with malaria
730 parasites.

731

732 For each blood smear image analysed, LIME produces a visual heatmap highlighting the most
733 influential regions in the classification decision. These highlighted areas allow healthcare
734 professionals to understand why the model classified an image as parasitised or uninfected.
735 This feature is crucial in clinical settings because it provides clinicians with tangible visual
736 cues to validate the model's predictions. It bridges the gap between complex machine learning
737 models and human interpretability, making AI-driven diagnostic systems more transparent and
738 trustworthy.

739

740 By offering healthcare professionals clear, visual evidence of what the model “sees,” LIME
741 enhances decision-making support. Physicians and laboratory technicians are given the output
742 of the model and the rationale behind the decision, building trust in the diagnostic process.

743 Moreover, the mathematical framework behind LIME ensures that the explanations are robust
744 and focused on the most significant aspects of the model's behaviour.

745

746 2.6.3 Integration with LLaMA for Diagnostic Insights

747

748 The above outputs were complemented with textual descriptions produced by the Large
749 Language Model Meta AI (LLaMA) language model to provide more specific case information
750 (87). It enabled the app to classify blood smear images and briefly describe each classification
751 based on context. An advantage of the system was that it used a pre-trained LLaMA model to
752 provide interpretations in simple language of why a particular classification of the image as
753 parasitised or uninfected was made.

754

755 In the case of blood smears classified as parasitised, the LLaMA model offered some
756 understanding of the visual cues that contributed to this classification. It also showed the
757 presence of trophozoites, the asexual form of malaria parasite presenting as ring-shaped
758 structures within red blood cells and some other abnormalities, including irregular shapes of
759 the red blood cells. When these features were identified by the ConvNeXt models, they marked
760 a critical point in signalling an infection, which LLaMA translated into a layman's
761 understanding.

762

763 The use of natural language processing in combination with machine learning algorithms has
764 its advantages in clinical settings as it helps make the results more understandable. This way,
765 the system expands the original AI's predictions with natural language descriptions to fulfil the
766 requirements of healthcare workers. Extending the ability to justify the prediction made by the
767 model helps enhance the credibility of the model's predictions. It enables the clinician to act

768 on the insights provided by the AI system more effectively. Ultimately, this integration makes
769 the system more practically beneficial for diagnostic pipelines.

770

771 2.6.4 Deployment

772

773 To develop the application, a Gradio interface with a well-organised and easily navigable
774 layout was employed to facilitate the uploading of microscopic blood smear images with the
775 subsequent immediate diagnostic output. When an image is uploaded, the system processes the
776 image through a pre-processing, classification, and explanation phase. The results are presented
777 to the user in both textual and graphical formats. It also builds on the features of Gradio's
778 interface that enhance the interaction, especially for individuals who may not be quite
779 conversant with the software. After the image has been uploaded, the fine-tuned ConvNeXt
780 models are applied to the image, and the output is either parasitised or uninfected. In addition,
781 a LIME model visualises the image regions that led to the classification decision.

782

783 The additional textual explanation that LLaMA can help practitioners understand the outcomes
784 more straightforwardly. This dual output, which is both visual and textual, significantly
785 improves the understandability of the AI-based differential diagnosis. For this reason, the app
786 has a supporting backend script that controls memory usage, especially in environments with
787 constrained resources. It is the same whether the code runs on a GPU for fast computations or
788 a CPU in less powerful environments; the app backend is designed to be non-memory bound.

789

790 This design helps the app work effectively even in regions with limited resources, which is
791 typical for malaria-endemic areas. Since the app was developed based on Gradio, it has a high
792 level of adaptability. It can be adapted for mobile versions or integrated into existing healthcare

793 systems. This portability is especially valuable in malaria-endemic areas where portable
794 diagnostic tools are crucial for front-line healthcare providers. With this app accessible online,
795 installed on mobile devices or in healthcare systems of a specific region, the lack of time and
796 AI in detecting malaria becomes relative, thus increasing the chances of curing patients in areas
797 where malaria still needs to be solved.

798

799 **3. Results**

800

801 In this study, deep learning has been widely used for malaria parasite detection, and many
802 evaluation metrics have been used to assess the performance of the models in each of the
803 classification tasks. To evaluate the performance of the selected models such as Swin Tiny,
804 ResNet18, ResNet50, ConvNeXt Tiny, ConvNeXt V2 Tiny and a re-modified version of
805 ConvNeXt V2 Tiny, various criteria were used. These metrics are accuracy, precision, recall,
806 F1 score and ROC-AUC (Receiver Operating Characteristic – Area Under the Curve). Relative
807 measures, including log loss, MCC, specificity, balanced accuracy, Cohen’s kappa, G-mean,
808 FPR and FNR, were also used to assess the discriminative ability of the models between
809 parasitised and uninfected blood smears.

810

811 The basic measure is accuracy, in Equation 10, and is calculated as the number of correctly
812 classified observations divided by the total observations. Mathematically, accuracy can be
813 expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

814 The symbols used are TP for True Positives, TN for True Negatives, FP for False Positives and
815 FN for False Negatives. Since the dataset used in this study was balanced, accuracy is a good

816 measure to use, although it may not be the best measure for imbalanced data. Both precision
817 and recall, thus, give a more detailed picture.

818

819 Precision or Positive Predictive Value, shown in Equation 11, is the probability that a positive
820 identification is correct. It is calculated as:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

821 Precision is crucial when false positives are to be avoided, for example, in a diagnosis where a
822 wrong result can mean a patient is subjected to a treatment he does not need.

823

824 Recall, or Sensitivity, is the ability to find all the positives out of the cases that are in fact
825 positive. It is calculated as:

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

826 A high recall rate, calculated as shown in Equation 12, means that the model accurately predicts
827 most of the positives, which is especially important for tasks that involve risk, such as
828 diagnosing malaria, when failure to identify a positive case (false negative) may have severe
829 outcomes.

830

831 The F1 score, in Equation 13, integrates precision and recall into a single score that considers
832 both of them equally profitable or unprofitable. The F1 score is defined as the harmonic mean
833 of precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

834 This metric gives us a more holistic picture of the model's performance when it needs to
835 balance trade-offs between precision and recall.

836

837 The ROC-AUC was also considered in this study, as shown in Equation 14. The true positive
838 rate (TPR) is plotted against the false positive rate (FPR), giving the ROC curve, which
839 graphically represents how the model is discriminatory. This performance is summarised by
840 the AUC (Area Under the Curve) where the AUC score of 1 represents perfect discrimination
841 and of 0.5 represents no discrimination. The AUC can be mathematically expressed as:

$$ROC - AUC = \int_0^1 TPR(FPR)d(FPR) \quad (14)$$

842 A particularly useful metric for model performance in medical diagnostics is ROC AUC, whose
843 purpose is to evaluate model performance across different thresholds and, thereby, across a
844 range of decision boundaries.

845

846 Furthermore, uncertainty of predictions was measured using log loss, in Equation 15. This
847 metric is a rigorous metric because it penalises confident incorrect predictions more heavily
848 than less confident ones. Mathematically, it is defined as:

$$Log\ loss = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (15)$$

849 where y_i is the actual label and p_i is the predicted probability for the positive class.

850

851 Quality of binary classifications was measured using Matthews correlation coefficient (referred
852 to as MCC), given as Equation 16. It covers all four confusion matrix categories (TP, TN, FP,
853 FN) and works well in imbalanced dataset cases, even though, the dataset used in this particular
854 case was balanced. MCC is calculated as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TN + FP)(TP + FN)}} \quad (16)$$

855 True Negative Rate, OR Specificity, is the number of actual negatives identified as negatives.

856 Recall addresses positive cases, and it is complemented by it. Agreement between the model's

857 predictions and true labels were also measured using Cohen's kappa, refined by chance, and
858 G-mean that is a geometric mean of sensitivity and specificity.

859

860 Moreover, using both False Positive Rate (FPR), in Equation 17, and False Negative Rate
861 (FNR), in Equation 18, gives additional information about what type of errors the models made.

862 FPR, calculated as:

$$FPR = \frac{FP}{FP + TN} \quad (17)$$

863 indicates how often the model incorrectly labels uninfected images as parasitised, whereas

864 FNR, calculated as:

$$FNR = \frac{FN}{FN + TP} \quad (18)$$

865 reflects how often the model fails to detect actual parasitic infections.

866

867 Every metric has its advantages and disadvantages. For instance, while accuracy is easy to
868 understand and calculate, it can be misleading with imbalanced datasets, though that was not
869 an issue in the dataset used in this study. Precision is relevant in situations with the least
870 tolerance for error in predictions. At the same time, recall is important in situations requiring
871 detecting as many positive cases as possible. The advantage of the F1 score is that it is a
872 balanced metric, but it can be less useful when precision and recall are much different. ROC-
873 AUC is a robust measure in different thresholds. However, it can be sensitive to the presence
874 of imbalanced data; this was not a problem in this work. The log loss ensures a model is not
875 overconfident in its predictions, and MCC provides a more complementary score for
876 performance, especially when dealing with imbalanced datasets. FNR and G-mean increase the
877 understanding of the model's failure modes. Tables 4.1, 4.2 and 4.3 present model performance
878 comparisons across these critical metrics.

879

880 Table 4.1: Model performance comparison across critical metrics

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Swin Tiny	0.613988	0.572842	0.896420	0.699000	0.679830
ResNet18	0.625687	0.572930	0.987379	0.725112	0.839727
ResNet50	0.813671	0.731378	0.991502	0.841803	0.951111
ConvNeXt Tiny	0.958646	0.944025	0.975110	0.959316	0.991124
ConvNeXt V2 Tiny	0.544330	0.523228	0.998578	0.686663	0.815418
ConvNeXt V2 Tiny Remod	0.981249	0.979453	0.983123	0.981285	0.996633

881

882 Table 4.2: Model performance comparison across critical metrics

Model	Log Loss	MCC	Specificity	Balanced Accuracy
Swin Tiny	1.402099	0.276272	0.331555	0.613988
ResNet18	1.681094	0.364074	0.263995	0.625687
ResNet50	0.831209	0.671230	0.635839	0.813671
ConvNeXt Tiny	0.116142	0.917790	0.942181	0.958646
ConvNeXt V2 Tiny	2.224143	0.212159	0.090081	0.544330
ConvNeXt V2 Tiny Remod	0.099783	0.962506	0.979376	0.981249

883

884 Table 4.3: Model performance comparison across critical metrics

Model	Cohen's Kappa	G-Mean	FPR	FNR
Swin Tiny	0.227975	0.545172	0.668445	0.103580
ResNet18	0.251374	0.510552	0.736005	0.012621
ResNet50	0.627341	0.794000	0.364161	0.008498
ConvNeXt Tiny	0.917292	0.958505	0.057819	0.024890
ConvNeXt V2 Tiny	0.088659	0.299922	0.909919	0.001422
ConvNeXt V2 Tiny Remod	0.962499	0.981248	0.020624	0.016877

885

886 The ConvNeXt models, especially the ConvNeXt V2 Tiny Remod, were highly able to
887 differentiate between parasitised and uninfected samples with an accuracy of 98.1%, as shown
888 in Table 4.1. This means the model can differentiate between malaria and other samples with
889 high accuracy and low chances of misclassification, making it suitable for real-world
890 applications. However, in the case of the Swin Tiny model, the accuracy was relatively low,
891 standing at 61.4%, significantly indicating some discrepancies in the classification when
892 differentiating between the two types of samples, which may raise a question regarding its
893 practicability for real-time diagnosis.

894

895 Another critical factor in measuring the models' performance is their capability to minimise
896 the number of false positives. Here, the accuracy of the ConvNeXt V2 Tiny Remod was
897 impressive, with a precision of 97.9%, as shown in Table 4.1. This high precision shows that
898 the model made only a few mistakes when identifying parasitised samples, thus reducing the
899 possibility of unnecessary treatment for unparasitised ones. On the other hand, the precision of

900 Swin Tiny stands at 57.3%, which is an indication of the model's propensity to classify healthy
901 samples as infected ones. Such high false positive readings can result in mistreating patients
902 and unnecessary procedures in a clinical context.

903

904 The ability of the model to correctly predict actual malaria cases is also important, as measured
905 by recall. ConvNeXt V2 Tiny Remod had a strong recall, achieving 98.3% in this task. A recall
906 of 98.3% means that the model failed to identify only 1.7% of the parasitised cases, which
907 shows that the model has high sensitivity in detecting malaria-infected blood smear image.
908 Nevertheless, a higher recall of 89.6% was observed from Swin Tiny, which shows that the
909 model correctly identified a large number of parasitised samples; however, it also had a low
910 precision, meaning that Swin Tiny often misclassified uninfected samples as parasitised.

911

912 The F1 score, which incorporates the trade-off between precision and recall, also confirmed
913 the superiority of ConvNeXt V2 Tiny Remod with a score of 98.1%. This balance shows the
914 model's general performance in the classification task; the model could classify most positive
915 cases without classifying many as negative. On the same note, the F1 score of Swin Tiny at
916 69.9% indicates the model's inability to balance between detecting true positives and
917 minimising false positives ideally, thus being less suitable for the fine-tuning task of malaria
918 detection.

919

920 The accuracy and precision of the ConvNeXt models are not the only factors that make them
921 stand out. The ROC-AUC value of the model that quantifies how well the model can
922 differentiate parasitised and uninfected samples was 0.996 for ConvNeXt V2 Tiny Remod.
923 This high value points to the model being well-equipped to distinguish between the two
924 categories. However, the ROC-AUC of 0.679 for Swin Tiny shows that it did not perform well

925 in this regard, which reaffirms the previous observation that it was not quite adept at making
926 the correct classifications.

927

928 In terms of confidence in the predictions of the model, the ConvNeXt V2 Tiny Remod was
929 once more the best amongst the models, with a low log loss of 0.099. This low score suggests
930 that the model made highly certain decisions, thus limiting the chances of misdiagnosis to the
931 minimum. On the other end of the spectrum, with a log loss of 1.40, the Swin Tiny failed to
932 distinguish between parasitised and uninfected image classes, thus playing into its subpar
933 performance.

934

935 Additionally, ConvNeXt V2 Tiny Remod fared well in the MCC with a score of 0.962, which
936 measures the overall performance of a model for both false positives and false negatives. This
937 score enhances the credibility of the model in consistently producing accurate outcomes. Swin
938 Tiny's MCC was also relatively poor at 0.276, indicating that it was unreliable and had a high
939 level of variability in classifying the samples. This was especially seen in the model's
940 specificity, which did not produce many false positive results. Based on the results, ConvNeXt
941 V2 Tiny Remod had a specificity of 97.9%. Thus, it rarely offered false positive classifications
942 of uninfected samples as parasitised, which is essential to avoid unnecessary treatments for
943 patients who do not have the disease.

944

945 The sensitivity for Swin Tiny was 33.2%, a lower value compared to the other algorithms; this
946 can be attributed to the model's tendency to classify uninfected cells as parasitic, thus reducing
947 the model's usefulness in clinical diagnosis. Regarding balanced accuracy, that is, the mean of
948 recall and specificity, ConvNeXt V2 Tiny Remod occupies the leading position with a score of
949 98.1%. This high value indicates the model's good performance on both the positive and

950 negative classes because it can distinguish between them and avoid over-prediction cases
951 where it may erroneously predict the image as having parasites when it does not. This is
952 because the balanced accuracy of Swin Tiny stands at a low 61.4%, which shows
953 comprehensive inefficiency regarding balance.

954

955 The Cohen's kappa scores of the ConvNeXt models also showed that the models made almost
956 accurate classifications of the images. The ConvNeXt V2 Tiny Remod achieved an accuracy
957 of 0.962, which shows a high correlation with the true labels; on the other hand, Swin Tiny,
958 with an accuracy of 0.228, shows a low correlation and is in line with the earlier findings that
959 it performed poorly.

960

961 Considering G-Mean, a measure combining recall and specificity, ConvNeXt V2 Tiny Remod
962 is the best, with a value of 0.981. This implies that the model could distinguish between the
963 parasitised images and, at the same time, minimise over-prediction. For instance, Swin Tiny,
964 with a G-Mean of 0.545, failed to strike this balance, affecting its performance.

965

966 These findings indicate that the ConvNeXt models are substantially better than the others for
967 detecting malaria parasites in microscopic images, especially the ConvNeXt V2 Tiny Remod
968 model. Such models provided high accuracy and precision and had low error rates, thus
969 showing indications of suitability for practical use. Conversely, Swin Tiny could have been
970 stronger in several aspects, limiting its capability to address this classification problem.

971

972 The heatmap in Figure 7 supports the tabulated results as they compare model performance
973 across multiple metrics. Of all the models, ConvNeXt V2 Tiny Remod and ConvNeXt Tiny
974 have the best accuracy, precision, recall, and F1 score, as well as low log loss, thereby proving

975 to be efficient in identifying parasitised and uninfected blood smears. This is supported by their
976 high specificity and moderate sensitivity, which enhances their performance. Conversely, Swin
977 Tiny and ConvNeXt V2 Tiny models show less stable results with lower precision and higher
978 false positive rates, thus weaker classification capabilities. The general comparison indicates
979 that ConvNeXt V2 Tiny Remod is the most viable model for malaria identification.

980

981 Figure 7: Heatmap of model performance metrics

982

983 This assessment further shows that the ConvNeXt models, especially the ConvNeXt V2 Tiny
984 Remod, are more effective for malaria parasite detection in the blood smear slides. These
985 models have high accuracy, precision, and balanced classification for medical imaging tasks
986 and, therefore, have a great potential for enhancing diagnostic accuracy.

987

988 **4. Discussion**

989

990 This study shows that ConvNeXt architecture has many benefits and helps identify malaria
991 parasites in thin blood smear images. The architectural design of ConvNeXt for high-resolution
992 images allows it to have detailed local information and overall contextual information in the
993 images. This two-fold function is beneficial in tasks involving basic and comprehensive
994 information, such as medical diagnosis, since the images' small details and overall picture must
995 be analysed.

996

997 There are several reasons for the enhanced performance of ConvNeXt and, especially, the
998 ConvNeXt V2 Tiny Remod model. First, the hierarchical feature extraction of the model can
999 capture details of malaria parasites, including their shape, texture, and internal structure, while

1000 also capturing the general composition of the blood smear. This is very important in medical
1001 imaging, where diagnosis accuracy may depend on the finer details based on the image
1002 displayed.

1003

1004 Although this study's dataset was from Bangladesh and thus restricts the geographic
1005 generalisability, data augmentation was applied to introduce variability. Future work should
1006 include datasets from other regions with different malaria strains to improve the model's
1007 robustness. Collectively, the model could be expanded through global collaborations to
1008 improve its performance across different geographic contexts.

1009

1010 Another consideration for AI-based diagnostic systems is the possibility of bias arising from
1011 geographically or demographically limited training data. The dataset of this study mainly
1012 comprised blood smear images exclusively from Bangladesh. It raised issues with genetic
1013 diversity biases, such as ethnic differences in blood types and malaria strains. There is the
1014 potential that deploying the model in various regions or with different populations can affect
1015 its performance. Future studies should use more diverse datasets collected from different parts
1016 of the world, from different blood types, and different parasite strains to increase the
1017 generalisability of the model and ameliorate this bias.

1018

1019 Transfer learning was employed in the development of models in order to enhance their
1020 performance. The ImageNet dataset used to pre-train the models in this work provided them
1021 with a good starting point for general visual features such as edges, textures and shapes. This
1022 helped reduce the requirement for a large amount of annotated malaria data and improved the
1023 learning on the task-specific dataset.

1024

1025 To increase the model's robustness to poor-quality images that are more likely to be
1026 encountered in low-resource settings, the model was trained and tested on images with noise,
1027 rotations, and varying brightness, using data augmentation techniques. Furthermore, using
1028 many data augmentation techniques increased the model's generalisation capability. This
1029 enhanced the model's stability and guaranteed that the model's performance could be optimal
1030 in various practical situations.

1031

1032 However, these methods have some drawbacks which have been identified to affect their
1033 effectiveness. A limitation is that these models require access to computational resources,
1034 which may be scarce in some environments, particularly in resource-limited settings. Using
1035 architectures such as Swin-transformers and ConvNeXt can be computationally intensive, and
1036 this poses a challenge in practical applications where technology and infrastructure may be
1037 inadequate. On the other hand, access to these methods as an online service, where the
1038 computation can be deferred to reputable online platforms such as Amazon web services and
1039 Google Cloud, can be used as an alternative to enable access regardless of locally used
1040 computational resources (92-94).

1041

1042 While data augmentation and transfer learning could solve some problems, the problem of
1043 dataset representativeness still exists. The balanced dataset used in this study may not
1044 completely reflect the variability that can be potentially observed in real-world clinical settings,
1045 like staining technique, sample quality, or parasite life cycle stages. Additionally, the
1046 ConvNext models used in this study are the less advanced versions, ConvNext Tiny and
1047 ConvNext V2 Tiny models, due to the unavailability of computational resources. Future work
1048 can involve training more advanced models, such as ConvNext V2 XLarge, on more diverse
1049 and large datasets to find a way to generalise better.

1050

1051 The results from this study reveal the ability of ConvNeXt models, particularly in relation to
1052 the high recall and F1 scores that characterise these models and which are particularly useful
1053 for identifying malaria parasites. However, the utilisation of transfer learning and augmentation
1054 has drawbacks, including sensitivity to certain data distributions. This was somewhat offset by
1055 incorporating regularisation techniques such as the dropout.

1056

1057 Furthermore, including a new malaria diagnostic application within this framework emphasises
1058 the clinical significance of this study. The app builds upon ConvNeXt and incorporates
1059 methods such as LIME to facilitate real-time explainable AI diagnosis to healthcare
1060 professionals. This tool is especially useful in areas with a scarcity of specialists who are
1061 usually required to administer some of these tests. However, the app's deployment might be
1062 restricted by the requirement for considerable computing ability and constant connectivity to
1063 the Internet, indicating more efficient field versions.

1064

1065 The ConvNeXt architecture shows clear clinical significance for malaria diagnosis in resource-
1066 limited environments. In high-burden areas, manual microscopy based on traditional methods
1067 is highly dependent upon skilled personnel and is prone to human error. An AI-based approach
1068 using ConvNeXt models to solve malaria detection has been proposed. It is automated and
1069 accurate in detecting malaria parasites and simultaneously tackles some of the key limitations
1070 of traditional approaches.

1071

1072 Additionally, the high accuracy and precision of ConvNeXt V2 Tiny Remod (98% accurate)
1073 demonstrate that deep learning-based systems are reliable for detecting malaria with accuracy
1074 similar to expert-level diagnosis. Such reliability may significantly alleviate the diagnostic

1075 workload of high-volume clinics and facilitate speeding up the diagnostic process, offering
1076 timely treatment and improving patient outcomes. Adding the layer of clinical significance to
1077 explainable AI techniques like LIME (for visual explanations) allows healthcare professionals
1078 to understand the AI model's decision-making process. This puts into place an interpretability
1079 which builds trust with clinicians so that system outputs are both accurate and transparent.

1080

1081 **5. Conclusion**

1082

1083 Deep learning models, especially the ConvNeXt-based malaria detection models proposed in
1084 this study, are promising for real-world application, especially in low-resource settings.
1085 Malaria remains a major disease burden, especially in developing countries where there is a
1086 shortage of expert clinicians and well-equipped laboratories. The ConvNeXt V2 Tiny Remod
1087 model presented in this study with an accuracy of 98% presents an efficient method of
1088 automating malaria diagnosis. The scientific rationale is that ConvNeXt can be used to identify
1089 malaria parasites in blood smears because the architecture enables the model to capture detailed
1090 and contextual information as a microscopist would while examining blood smears under a
1091 microscope. This makes it very sensitive and specific in detection, which is vital in the early
1092 stages and correct identification of the parasites.

1093

1094 Moreover, data augmentation and transfer learning are implemented in the model, which
1095 improves the model's performance regardless of the imaging conditions. These models could
1096 help decrease diagnostic errors, reduce the time to diagnosis, and enhance the clinical
1097 management of malaria patients. The use of diagnostic services in mobile or edge devices also
1098 presents a way of extending the services to regions without access to such services, thus
1099 contributing to efforts to fight this global disease burden.

1100

1101 6. Bibliography

1102

- 1103 1. WHO. World malaria report 2021. Geneva: World Health Organization; 2021.
- 1104 2. Ullah H, Khan MI, Suleman NM, Ismail N, Khan Z, Sayyid G. A Review on Malarial
1105 Parasite. World Journal of Zoology. 2015;10(4):285-90.
- 1106 3. Makler MT, Palmer CJ, Ager AL. A review of practical techniques for the diagnosis of
1107 malaria. Annals of tropical medicine and parasitology. 1998;92(4):419-34.
- 1108 4. Ba EH, Baird JK, Barnwell J, Bell D, Carter J, Dhorda M, et al. Microscopy for the
1109 detection, identification and quantification of malaria parasites on stained thick and thin blood
1110 films in research settings: procedure: methods manual. 2015.
- 1111 5. Maturana CR, De Oliveira AD, Nadal S, Bilalli B, Serrat FZ, Soley ME, et al. Advances
1112 and challenges in automated malaria diagnosis using digital microscopy imaging with artificial
1113 intelligence tools: A review. Frontiers in microbiology. 2022;13:1006659.
- 1114 6. Yang R, Yu Y. Artificial convolutional neural network in object detection and semantic
1115 segmentation for medical imaging analysis. Frontiers in oncology. 2021;11:638182.
- 1116 7. Sadek FM, Solihin MI, Heltha F, Hong LW, Rizon M. A Comparison of Machine
1117 Learning and Deep Learning in Hyperspectral Image Classification. Enabling Industry 40
1118 through Advances in Mechatronics: Selected Articles from iM3F 2021, Malaysia: Springer;
1119 2022. p. 221-35.
- 1120 8. Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S, editors. A convnet for the
1121 2020s. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition;
1122 2022.
- 1123 9. Todi A, Narula N, Sharma M, Gupta U, editors. ConvNext: A Contemporary
1124 Architecture for Convolutional Neural Networks for Image Classification. 2023 3rd
1125 International Conference on Innovative Sustainable Computational Technologies (CISCT);
1126 2023: IEEE.
- 1127 10. Medicine NLo. NLM - Malaria Data. 2018.
- 1128 11. Yu H, Yang F, Rajaraman S, Ersoy I, Moallem G, Poostchi M, et al. Malaria Screener:
1129 a smartphone application for automated malaria screening. BMC Infectious Diseases.
1130 2020;20:1-8.
- 1131 12. Dablain D, Jacobson KN, Bellinger C, Roberts M, Chawla NV. Understanding CNN
1132 fragility when learning with imbalanced data. Machine Learning. 2024;113(7):4785-810.
- 1133 13. Kumar T, Verma K. A Theory Based on Conversion of RGB image to Gray image.
1134 International Journal of Computer Applications. 2010;7(2):7-10.
- 1135 14. Díaz G, Gonzalez F, Romero E, editors. Infected cell identification in thin blood images
1136 based on color pixel classification: comparison and analysis. Progress in Pattern Recognition,
1137 Image Analysis and Applications: 12th Iberoamericann Congress on Pattern Recognition,
1138 CIARP 2007, Valparaiso, Chile, November 13-16, 2007 Proceedings 12; 2007: Springer.
- 1139 15. Fatima T, Farid MS. Automatic detection of Plasmodium parasites from microscopic
1140 blood images. Journal of Parasitic Diseases. 2020;44(1):69-78.
- 1141 16. Hung Y-W, Wang C-L, Wang C-M, Chan Y-K, Tseng L-Y, Lee C-W, et al. Parasite
1142 and infected-erythrocyte image segmentation in stained blood smears. Journal of Medical and
1143 Biological Engineering. 2015;35:803-15.
- 1144 17. Bagui O, Zoueu J, Wählby C. Automatic malaria diagnosis by the use of multispectral
1145 contrast imaging. Journal of Physical Chemical News. 2015;75:86-98.

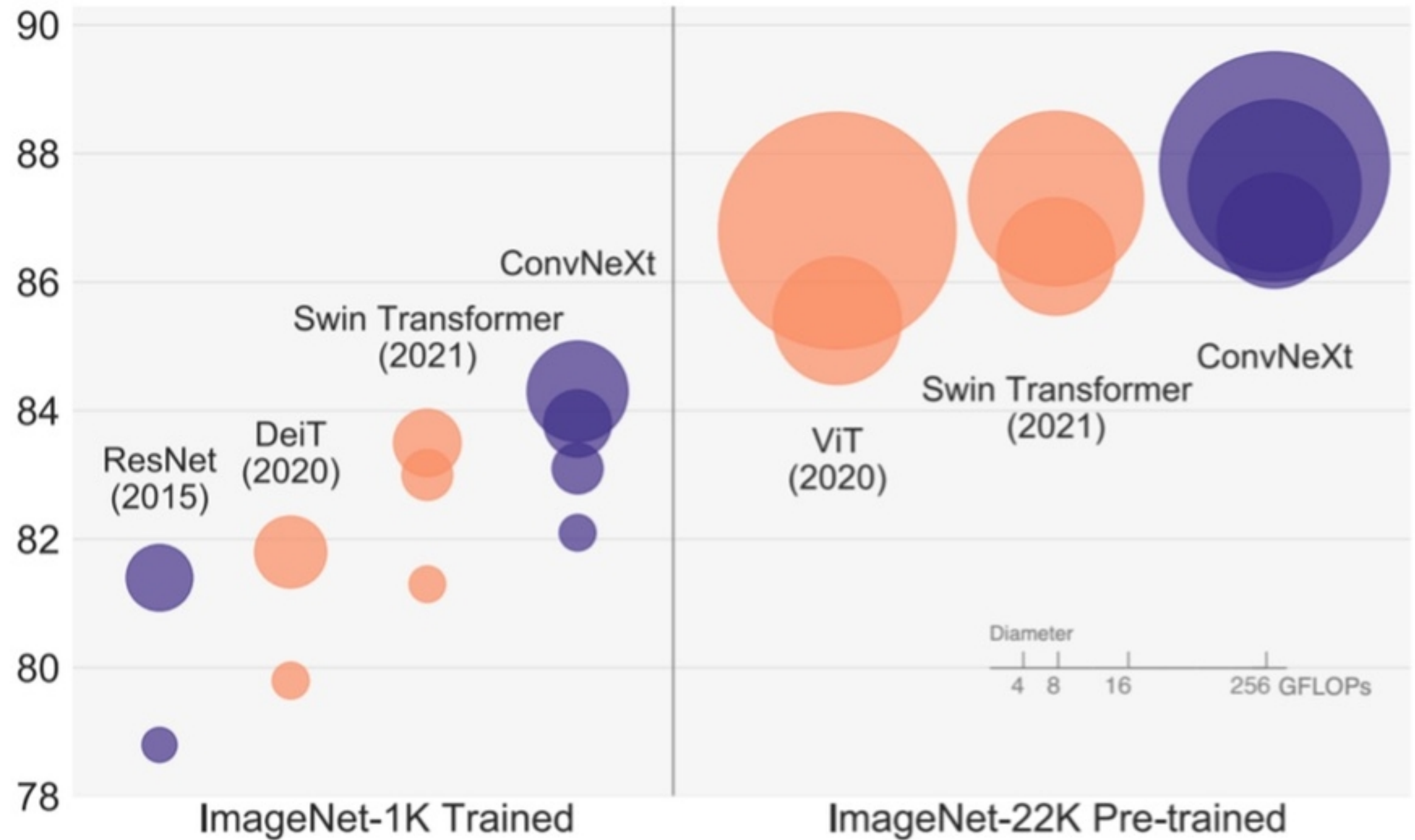
- 1146 18. Sharma H, Jain S, Vasudeva A. Detection of Malarial Parasite in Blood using Image
1147 Processing. 2023.
- 1148 19. Aksoy S, Demircioglu P, Bogrekci I. Enhancing Melanoma Diagnosis with Advanced
1149 Deep Learning Models Focusing on Vision Transformer, Swin Transformer, and ConvNeXt.
1150 Dermatopathology. 2024;11(3):239-52.
- 1151 20. Saponara S, Elhanashi A, editors. Impact of image resizing on deep learning detectors
1152 for training time and model performance. International Conference on Applications in
1153 Electronics Pervading Industry, Environment and Society; 2021: Springer.
- 1154 21. Moraes T, Amorim P, Da Silva JV, Pedrini H. Medical image interpolation based on
1155 3D Lanczos filtering. Computer Methods in Biomechanics and Biomedical Engineering:
1156 Imaging & Visualization. 2020;8(3):294-300.
- 1157 22. Magnusson A, Sandahl A. Artifact-Free Image Upscaling: An Evaluation of an
1158 Artifact-Free Color Interpolation Algorithm with Respect to Visual Quality. 2022.
- 1159 23. Madhukar B, Narendra R, editors. Lanczos resampling for the digital processing of
1160 remotely sensed images. Proceedings of International Conference on VLSI, Communication,
1161 Advanced Devices, Signals & Systems and Networking (VCASAN-2013); 2013: Springer.
- 1162 24. Yaroslavsky L. Fast discrete sinc-interpolation: a gold standard for image resampling.
1163 Advances in Signal Transforms: Theory and Applications. 2007;7:337-405.
- 1164 25. Bernstein GM, Gruen D. Resampling images in Fourier domain. Publications of the
1165 Astronomical Society of the Pacific. 2014;126(937):287.
- 1166 26. Thévenaz P, Blu T, Unser M. Image interpolation and resampling. Handbook of
1167 medical imaging, processing and analysis. 2000;1(1):393-420.
- 1168 27. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning (still)
1169 requires rethinking generalization. Communications of the ACM. 2021;64(3):107-15.
- 1170 28. Zheng S, Song Y, Leung T, Goodfellow I, editors. Improving the robustness of deep
1171 neural networks via stability training. Proceedings of the IEEE conference on computer vision
1172 and pattern recognition; 2016.
- 1173 29. Rahman A, Zunair H, Rahman MS, Yuki JQ, Biswas S, Alam MA, et al. Improving
1174 malaria parasite detection from red blood cell using deep convolutional neural networks. arXiv
1175 preprint arXiv:190710418. 2019.
- 1176 30. Fayaz S, Ahmad Shah SZ, Gul N, Assad A. Advancements in Data Augmentation and
1177 Transfer Learning: A Comprehensive Survey to Address Data Scarcity Challenges. Recent
1178 Advances in Computer Science and Communications (Formerly: Recent Patents on Computer
1179 Science). 2024;17(8):14-35.
- 1180 31. Mikołajczyk A, Grochowski M, editors. Data augmentation for improving deep
1181 learning in image classification problem. 2018 international interdisciplinary PhD workshop
1182 (IIPhDW); 2018: IEEE.
- 1183 32. Chandola Y, Uniyal V, Bachheti Y, Lakhera N, Rawat R. Data Augmentation
1184 Techniques applied to Medical Images. International Journal of Research Publication and
1185 Reviews. 2024;Vol 5, no 7:483-501.
- 1186 33. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning.
1187 Journal of big data. 2019;6(1):1-48.
- 1188 34. Hoorali F, Khosravi H, Moradi B. Automatic Bacillus anthracis bacteria detection and
1189 segmentation in microscopic images using UNet++. Journal of Microbiological Methods.
1190 2020;177:106056.
- 1191 35. Trung TQ, Bag A, Huyen LTN, Meeseepong M, Lee NE. Bio-Inspired Artificial
1192 Retinas Based on a Fibrous Inorganic–Organic Heterostructure for Neuromorphic Vision.
1193 Advanced Functional Materials. 2024;34(11):2309378.

- 1194 36. Shwetha V, Prasad K, Mukhopadhyay C, Banerjee B. Data augmentation for Gram-
1195 stain images based on Vector Quantized Variational AutoEncoder. *Neurocomputing*.
1196 2024;600:128123.
- 1197 37. Tiwari N, Omar M, Ghadi Y. Brain Tumor Classification From Magnetic Resonance
1198 Imaging Using Deep Learning and Novel Data Augmentation. *Transformational Interventions*
1199 *for Business, Technology, and Healthcare: IGI Global*; 2023. p. 392-413.
- 1200 38. Singh A, Bay A, Mirabile A. Assessing the importance of colours for cnns in object
1201 recognition. *arXiv preprint arXiv:201206917*. 2020.
- 1202 39. Abdollahi B, Tomita N, Hassanpour S. Data augmentation in training deep learning
1203 models for medical image analysis. *Deep learners and deep learner descriptors for medical*
1204 *applications*. 2020:167-80.
- 1205 40. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of
1206 medical image data augmentation techniques for deep learning applications. *Journal of Medical*
1207 *Imaging and Radiation Oncology*. 2021;65(5):545-63.
- 1208 41. Kim Y, Uddin AS, Bae S-H. Local augment: Utilizing local bias property of
1209 convolutional neural networks for data augmentation. *IEEE Access*. 2021;9:15191-9.
- 1210 42. Zhao J, Lu D, Ma K, Zhang Y, Zheng Y, editors. Deep image clustering with category-
1211 style representation. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK,*
1212 *August 23–28, 2020, Proceedings, Part XIV 16*; 2020: Springer.
- 1213 43. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: Hierarchical
1214 Vision Transformer using Shifted Windows. *Conference: 2021 IEEE/CVF International*
1215 *Conference on Computer Vision (ICCV)*; Montreal, QC, Canada: Institute of Electrical and
1216 *Electronics Engineers (IEEE)*; 2021.
- 1217 44. Tang Y, Yang D, Li W, Roth HR, Landman B, Xu D, et al. Self-Supervised Pre-
1218 Training of Swin Transformers for 3D Medical Image Analysis. *Conference: 2022 IEEE/CVF*
1219 *Conference on Computer Vision and Pattern Recognition (CVPR)*; New Orleans,
1220 Louisiana2022.
- 1221 45. Chen X, Qin Y, Xu W, Bur AM, Zhong C, Wang G, editors. Improving vision
1222 transformers on small datasets by increasing input information density in frequency domain.
1223 *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*; 2022.
- 1224 46. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition.
1225 *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016.
- 1226 47. Teow YJ. Malaria parasite detection from human blood smear images using deep
1227 learning techniques: UTAR; 2023.
- 1228 48. Cheng W, Liu J, Wang C, Jiang R, Jiang M, Kong F. Application of image recognition
1229 technology in pathological diagnosis of blood smears. *Clinical and Experimental Medicine*.
1230 2024;24(1):181.
- 1231 49. Li J, Wang C, Huang B, Zhou Z. ConvNeXt-backbone HoVerNet for nuclei
1232 segmentation and classification. *arXiv preprint arXiv:220213560*. 2022.
- 1233 50. Woo S, Debnath S, Hu R, Chen X, Liu Z, Kweon IS, et al., editors. Convnext v2: Co-
1234 designing and scaling convnets with masked autoencoders. *Proceedings of the IEEE/CVF*
1235 *Conference on Computer Vision and Pattern Recognition*; 2023.
- 1236 51. Müller R, Kornblith S, Hinton GE. When does label smoothing help? *Advances in*
1237 *neural information processing systems*. 2019;32.
- 1238 52. Lukasik M, Bhojanapalli S, Menon A, Kumar S, editors. Does label smoothing mitigate
1239 label noise? *International Conference on Machine Learning*; 2020: PMLR.
- 1240 53. Reyad M, Sarhan AM, Arafa M. A modified Adam algorithm for deep neural network
1241 optimization. *Neural Computing and Applications*. 2023;35(23):17095-112.

- 1242 54. Chen X, Liang C, Huang D, Real E, Liu Y, Wang K, et al., editors. Evolved optimizer
1243 for vision. First Conference on Automated Machine Learning (Late-Breaking Workshop);
1244 2022.
- 1245 55. Meng LK, Xin LJ, Yi HH, Salam ZAA, Wei NB. A machine learning approach for face
1246 mask detection system with AdamW optimizer. *J Appl Technol Innov.* 2023;7(3):25.
- 1247 56. Omidvar S, Tran T. Tackling cold-start with deep personalized transfer of user
1248 preferences for cross-domain recommendation. *International Journal of Data Science and*
1249 *Analytics.* 2023:1-10.
- 1250 57. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer
1251 learning. *Proceedings of the IEEE.* 2020;109(1):43-76.
- 1252 58. Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *Journal of Big*
1253 *data.* 2016;3:1-40.
- 1254 59. Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: Understanding transfer
1255 learning for medical imaging. *Advances in neural information processing systems.* 2019;32.
- 1256 60. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and*
1257 *data engineering.* 2009;22(10):1345-59.
- 1258 61. Liu ZaL, Yutong and Cao, Yue and Hu, Han and Wei, Yixuan and Zhang, Zheng and
1259 Lin, Stephen and Guo, Baining. Swin Transformer: Hierarchical Vision Transformer using
1260 Shifted Windows *Proceedings of the IEEE/CVF International Conference on Computer Vision*
1261 *(ICCV)2021* [
1262 62. Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, et al. A survey on vision transformer.
1263 *IEEE transactions on pattern analysis and machine intelligence.* 2022;45(1):87-110.
- 1264 63. Zhuang Liu HM, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining
1265 Xie. A ConvNet for the 2020s 2022 [Available from:
1266 <https://github.com/facebookresearch/ConvNeXt?tab=readme-ov-file>.
1267 64. Sanghyun Woo SD, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, Saining
1268 Xie. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders 2023
1269 [Available from: <https://github.com/facebookresearch/ConvNeXt-V2>.
1270 65. Marcelino P. Transfer learning from pre-trained models. *Towards data science.*
1271 2018;10(330):23.
- 1272 66. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional
1273 neural networks. *Advances in neural information processing systems.* 2012;25.
- 1274 67. Montalbo FJP, Alon AS. Empirical analysis of a fine-tuned deep convolutional model
1275 in classifying and detecting malaria parasites from blood smears. *KSII Transactions on Internet*
1276 *and Information Systems (TIIS).* 2021;15(1):147-65.
- 1277 68. Dörrich M, Fan M, Kist AM. Impact of Mixed Precision Techniques on Training and
1278 Inference Efficiency of Deep Neural Networks. *IEEE Access.* 2023;11:57627-34.
- 1279 69. Nokhwal S, Chilakalapudi P, Donekal P, Nokhwal S, Pahune S, Chaudhary A, editors.
1280 *Accelerating neural network training: A brief review. Proceedings of the 2024 8th International*
1281 *Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence; 2024.*
- 1282 70. Kandel I, Castelli M, Popovič A. Comparative study of first order optimizers for image
1283 classification using convolutional neural networks on histopathology images. *Journal of*
1284 *imaging.* 2020;6(9):92.
- 1285 71. Okewu E, Misra S, Lius F-S, editors. Parameter tuning using adaptive moment
1286 estimation in deep learning neural networks. *Computational Science and Its Applications–*
1287 *ICCSA 2020: 20th International Conference, Cagliari, Italy, July 1–4, 2020, Proceedings, Part*
1288 *VI 20; 2020: Springer.*
- 1289 72. Lee HH. Exploring Explainable Optimization in Medical Segmentation Network for
1290 Multi-Scale Generalization With Anatomical Atlas: Vanderbilt University; 2023.
- 1291 73. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature.* 2015;521(7553):436-44.

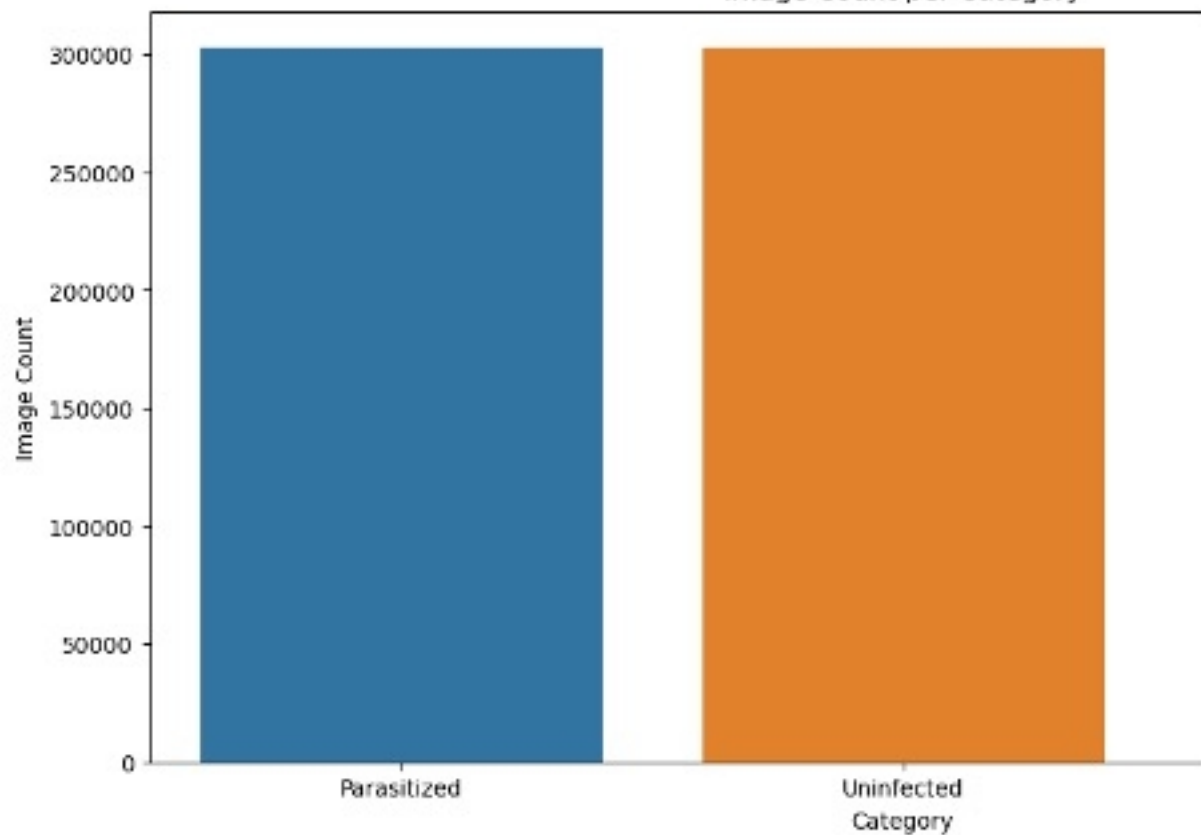
- 1292 74. Li S, Liu Z, Tian J, Wang G, Wang Z, Jin W, et al. Switch EMA: A Free Lunch for
1293 Better Flatness and Sharpness. arXiv preprint arXiv:240209240. 2024.
- 1294 75. Pan Y, Li Y. Toward understanding why adam converges faster than sgd for
1295 transformers. arXiv preprint arXiv:230600204. 2023.
- 1296 76. Horváth S, Mishchenko K, Richtárik P. Adaptive learning rates for faster stochastic
1297 gradient methods. arXiv preprint arXiv:220805287. 2022.
- 1298 77. Zhao R, Vogel B, Ahmed T, Luk W, editors. Reducing underflow in mixed precision
1299 training by gradient scaling. Proceedings of the Twenty-Ninth International Conference on
1300 International Joint Conferences on Artificial Intelligence; 2021.
- 1301 78. Kim C, Kim S, Kim J, Lee D, Kim S. Automated learning rate scheduler for large-batch
1302 training. arXiv preprint arXiv:210705855. 2021.
- 1303 79. Carvalho P, Lourenço N, Assunção F, Machado P, editors. Autolr: An evolutionary
1304 approach to learning rate policies. Proceedings of the 2020 genetic and evolutionary
1305 computation conference; 2020.
- 1306 80. Ruby U, Yendapalli V. Binary cross entropy with deep learning technique for image
1307 classification. Int J Adv Trends Comput Sci Eng. 2020;9(10).
- 1308 81. Ho Y, Wookey S. The real-world-weight cross-entropy loss function: Modeling the
1309 costs of mislabeling. IEEE access. 2019;8:4806-13.
- 1310 82. Wang Y, Ma X, Chen Z, Luo Y, Yi J, Bailey J, editors. Symmetric cross entropy for
1311 robust learning with noisy labels. Proceedings of the IEEE/CVF international conference on
1312 computer vision; 2019.
- 1313 83. Mmileng OP. Malaria Diagnosis App 1.0 Johannesburg2024 [Available from:
1314 <https://huggingface.co/spaces/Phikhei/ConvNextMalariaDetector>].
- 1315 84. Garreau D, Luxburg U, editors. Explaining the explainer: A first theoretical analysis of
1316 LIME. International conference on artificial intelligence and statistics; 2020: PMLR.
- 1317 85. Aldughayfiq B, Ashfaq F, Jhanjhi N, Humayun M. Explainable AI for retinoblastoma
1318 diagnosis: interpreting deep learning models with LIME and SHAP. Diagnostics.
1319 2023;13(11):1932.
- 1320 86. Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, Letman A, et al. The llama 3
1321 herd of models. arXiv preprint arXiv:240721783. 2024.
- 1322 87. Vavekanand R, Sam K. Llama 3.1: An In-Depth Analysis of the Next-Generation Large
1323 Language Model. ResearchGate; 2024.
- 1324 88. Rane N, Choudhary SP, Rane J. Ensemble deep learning and machine learning:
1325 applications, opportunities, challenges, and future directions. Studies in Medical and Health
1326 Sciences. 2024;1(2):18-41.
- 1327 89. Huff DT, Weisman AJ, Jeraj R. Interpretation and visualization techniques for deep
1328 learning models in medical imaging. Physics in Medicine & Biology. 2021;66(4):04TR1.
- 1329 90. Mahmood Y, Kama N, Azmi A, Ali M, editors. Improving estimation accuracy
1330 prediction of software development effort: A proposed ensemble model. 2020 International
1331 Conference on Electrical, Communication, and Computer Engineering (ICECCE); 2020: IEEE.
- 1332 91. Etemadi S, Khashei M. Accuracy versus reliability-based modelling approaches for
1333 medical decision making. Computers in Biology and Medicine. 2022;141:105138.
- 1334 92. Li A, Yang X, Kandula S, Zhang M, editors. CloudCmp: comparing public cloud
1335 providers. Proceedings of the 10th ACM SIGCOMM conference on Internet measurement;
1336 2010.
- 1337 93. Srinivasan V, Ravi J, Raj J. Google Cloud Platform for Architects: Design and manage
1338 powerful cloud solutions: Packt Publishing Ltd; 2018.
- 1339 94. Reese G. Cloud application architectures: building applications and infrastructure in the
1340 cloud: " O'Reilly Media, Inc."; 2009.
- 1341

ImageNet-1K Acc.

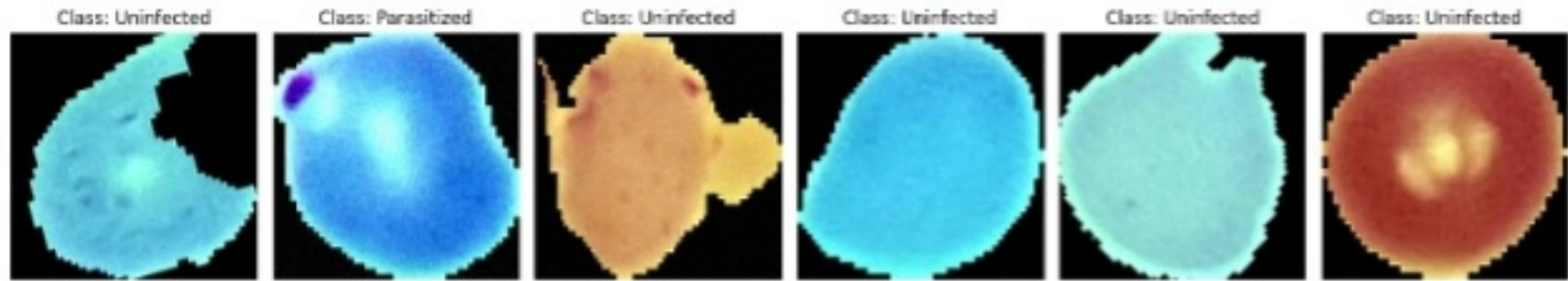


Figure

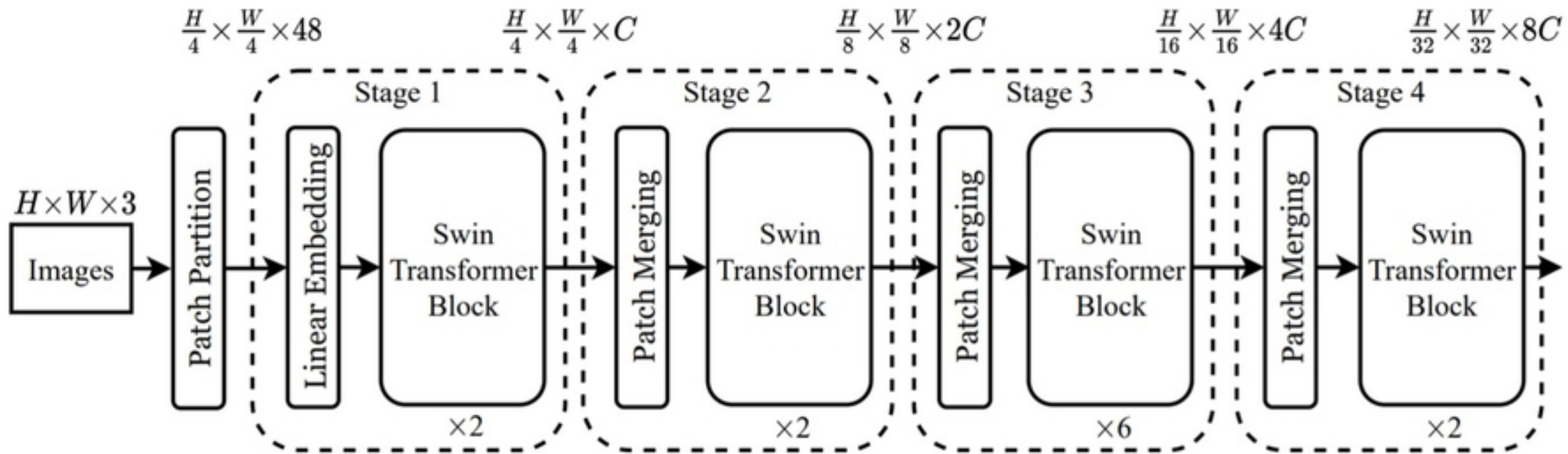
Image Count per Category



Figure

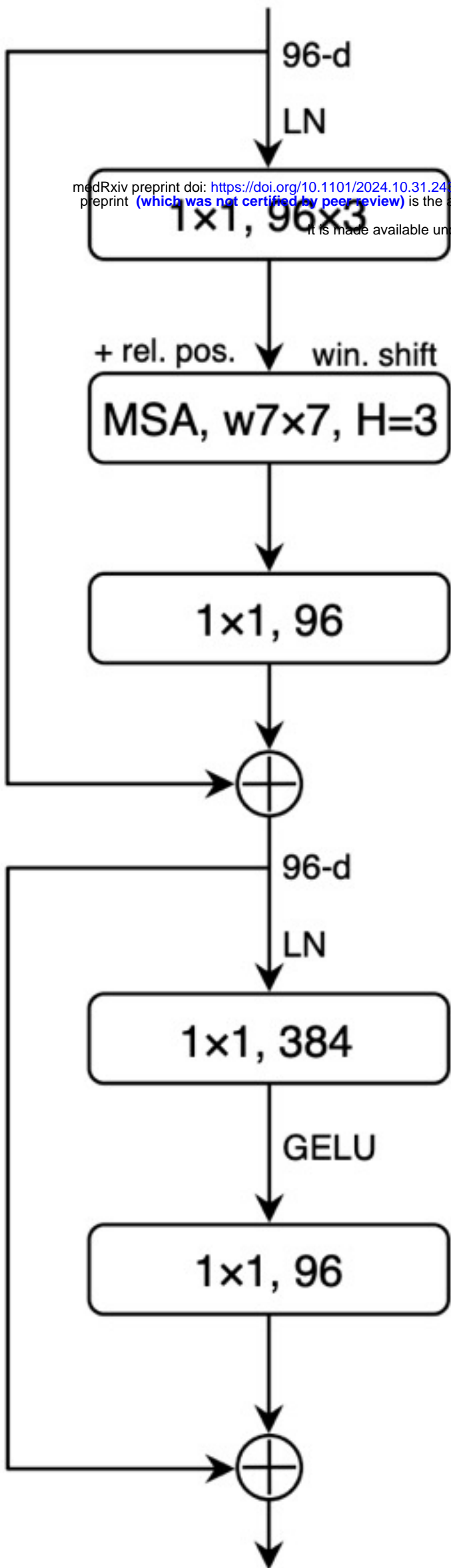


Figure



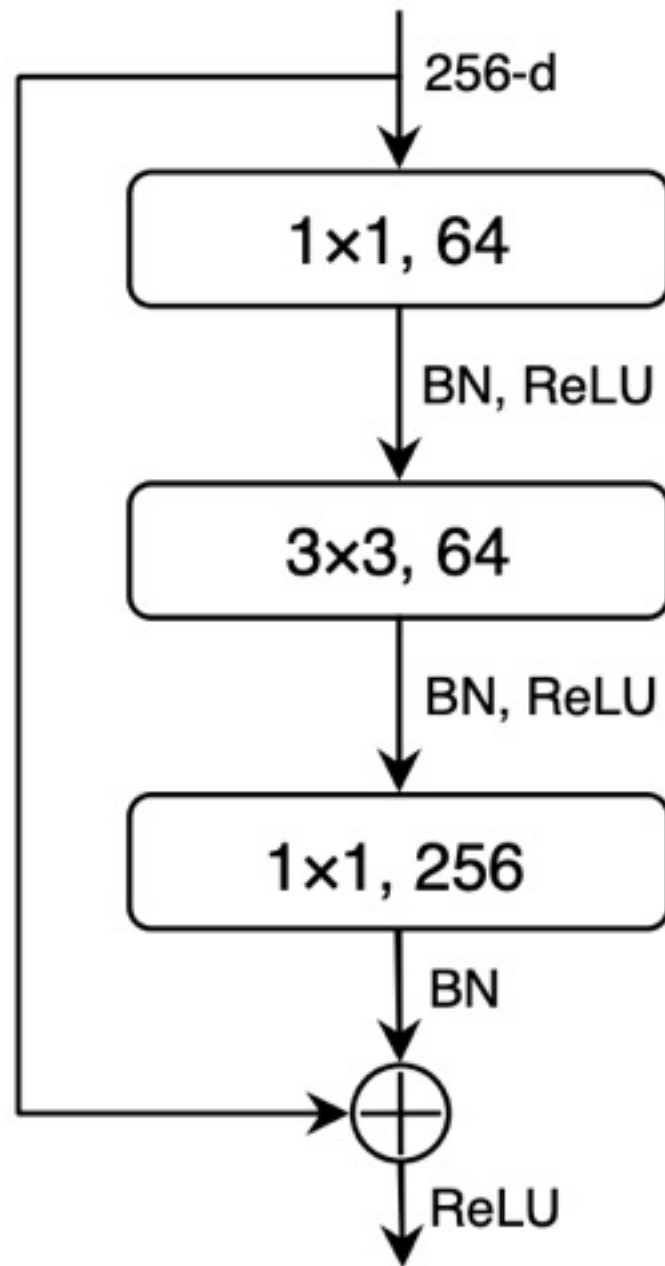
Figure

Swin Transformer Block

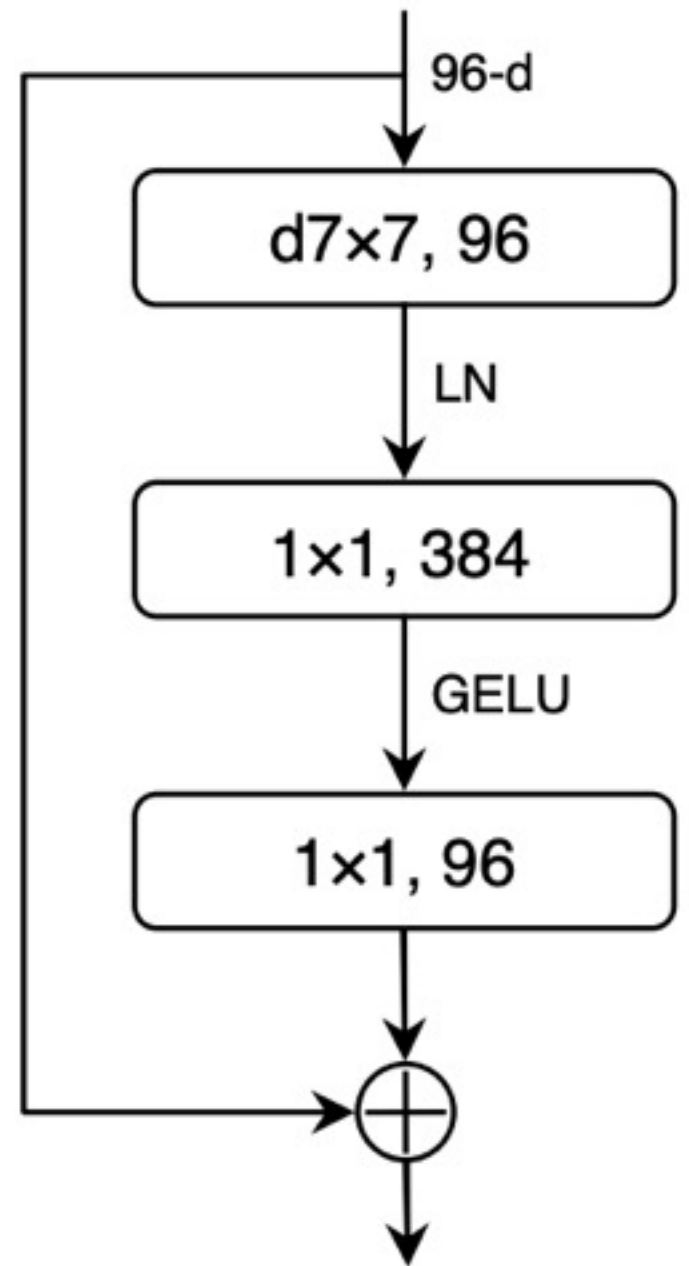


medRxiv preprint doi: <https://doi.org/10.1101/2024.10.31.24316549>; this version posted November 4, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

ResNet Block

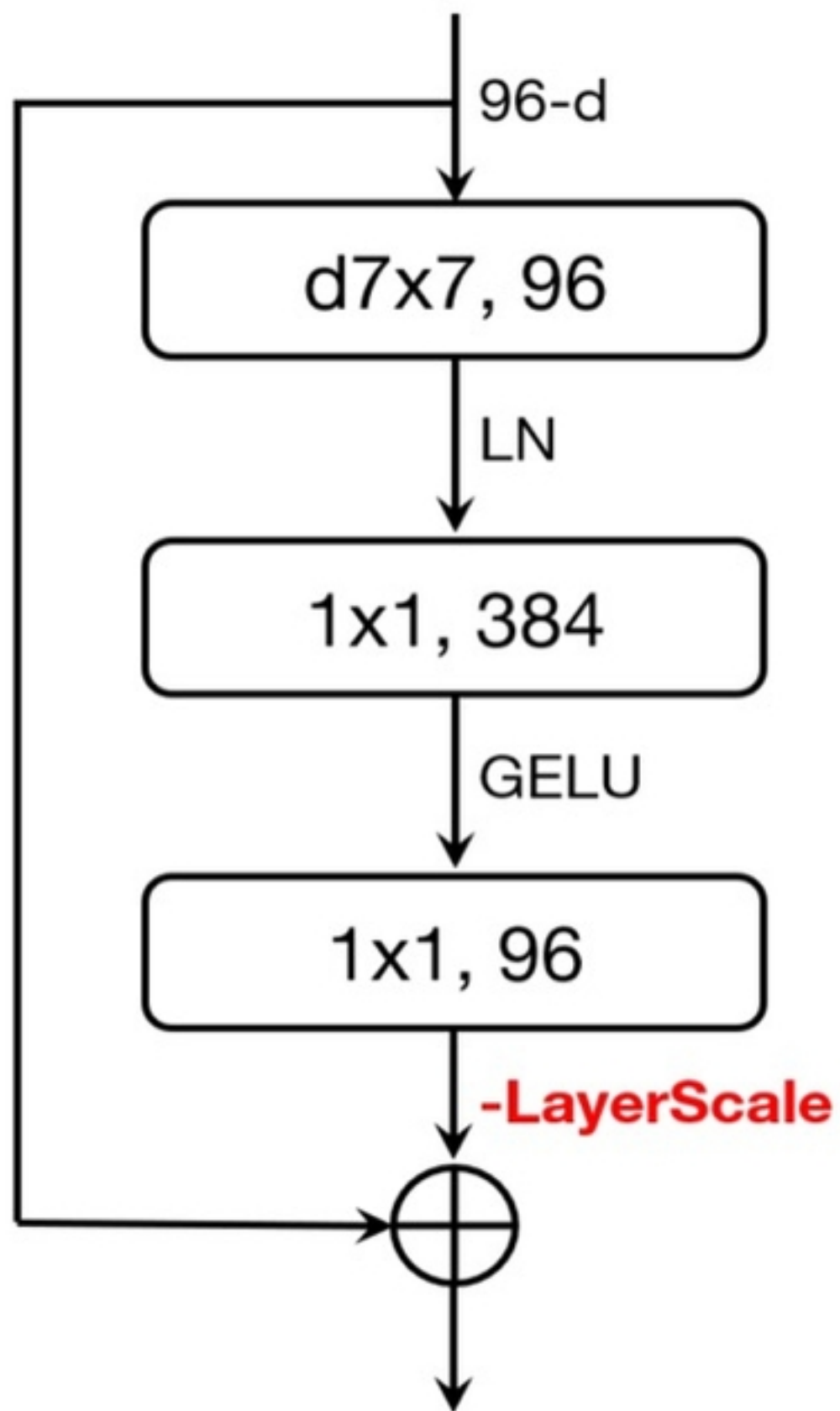


ConvNeXt Block

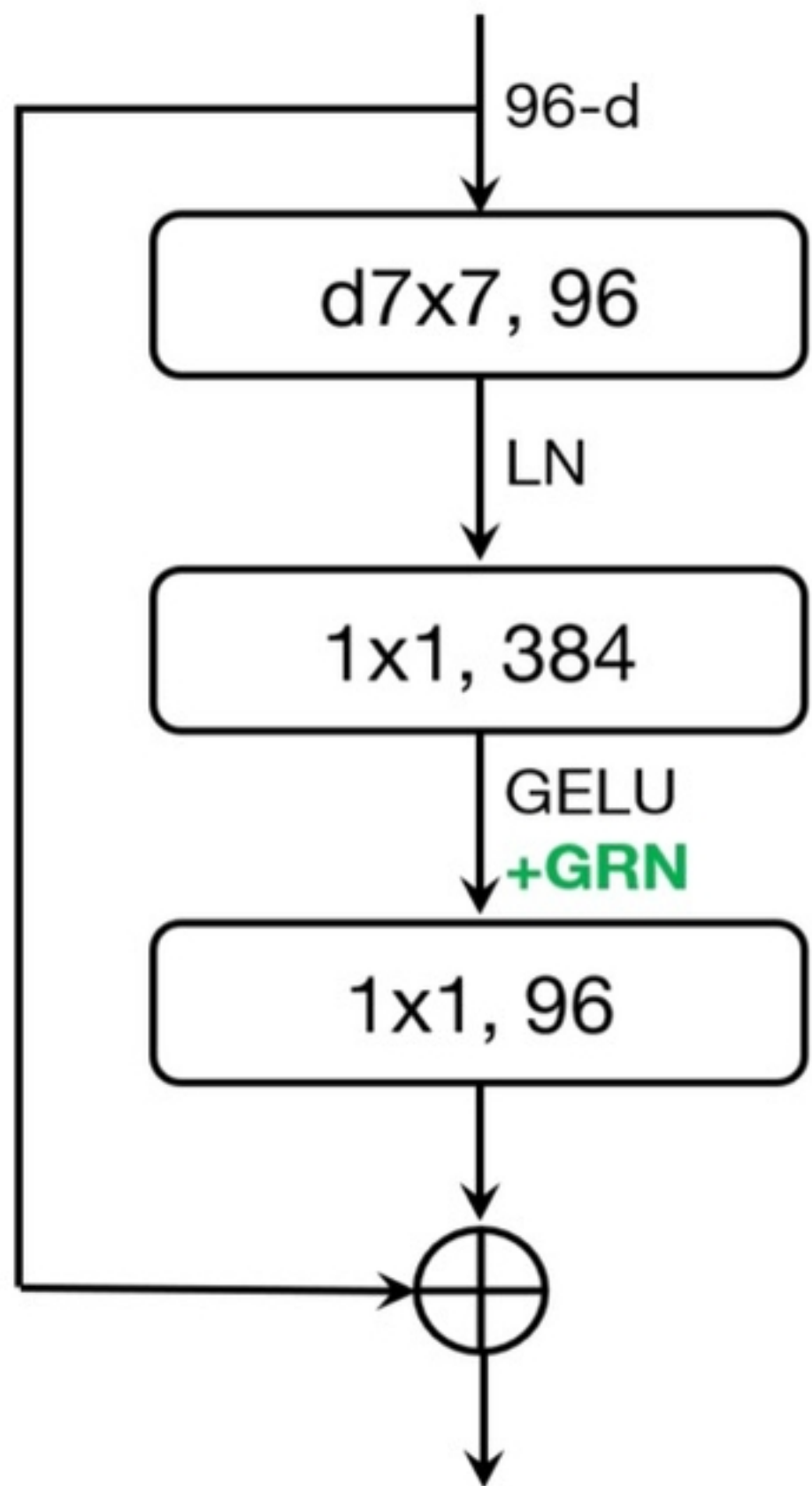


Figure

ConvNeXt V1 Block

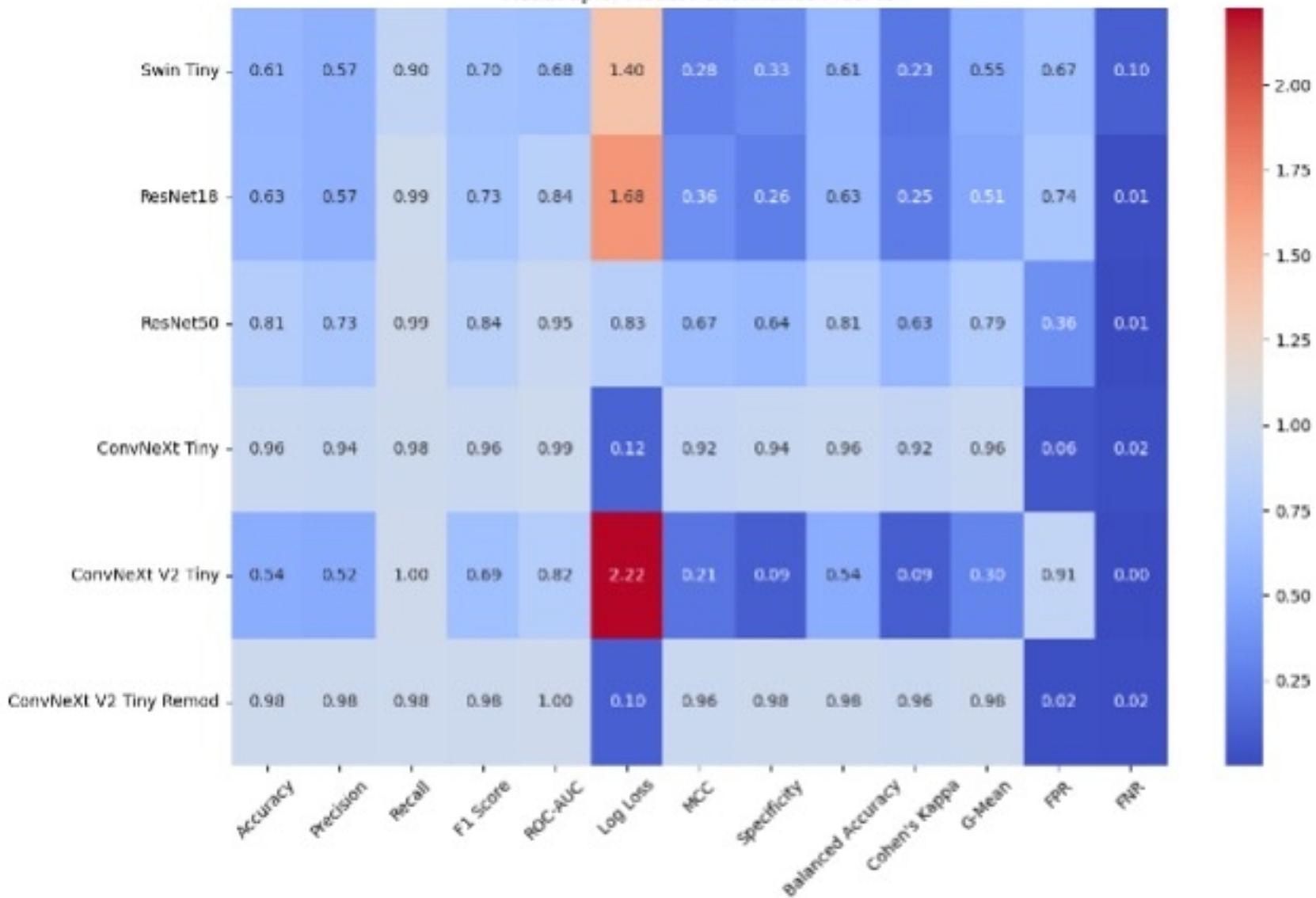


ConvNeXt V2 Block



Figure

Heatmap of Model Performance Metrics



Figure