

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Street images classification according to COVID-19 risk in Lima, Peru: A convolutional neural networks feasibility analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-063411
Article Type:	Original research
Date Submitted by the Author:	30-Mar-2022
Complete List of Authors:	Carrillo-Larco, Rodrigo; Imperial College London, Castillo-Cara, Manuel; Universidad Politécnica de Madrid, Ontology Engineering Group Hernández Santa Cruz, Jose Francisco; Independent Researcher
Keywords:	COVID-19, EPIDEMIOLOGY, PUBLIC HEALTH

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3
4
5 1 **Street images classification according to COVID-19 risk in Lima, Peru: A convolutional neural**
6
7 2 **networks feasibility analysis**
8
9 3

10
11 4 Rodrigo M Carrillo-Larco^{1,2,3}
12
13

14 5 Manuel Castillo-Cara^{4,5}
15
16

17 6 Jose Francisco Hernández Santa Cruz⁶
18
19

- 20 7
21
22 8 1. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,
23 London, UK
24 9
25 10 2. CRONICAS Centre of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia,
26 Lima, Peru
27 11
28 12 3. Universidad Continental, Lima, Peru
29 13
30 14 4. Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
31 15
32 16 5. Universidad de Lima, Lima, Peru
33 17
34 18 6. Independent researcher, Edinburgh, UK
35 19
36 20
37 21
38 22
39 23
40 24
41 25
42 26
43 27
44 28
45 29
46 30
47 31
48 32
49 33
50 34
51 35
52 36
53 37
54 38
55 39
56 40
57 41
58 42
59 43
60 44

41 17 **Corresponding author:**
42
43

44 18 Rodrigo M Carrillo-Larco, MD
45

46 19 Department of Epidemiology and Biostatistics
47

48 20 School of Public Health
49

50 21 Imperial College London
51

52 22 rcarrill@imperial.ac.uk
53
54
55
56
57
58
59
60

23 ABSTRACT

24 **Objectives:** During the COVID-19 pandemic convolutional neural networks (CNNs) have been used in
25 clinical medicine (e.g., chest X-rays classification). Whether CNNs could inform the epidemiology of COVID-
26 19 classifying street images according to COVID-19 risk is unknown, yet it could pinpoint high-risk places
27 and relevant features of the built environment. In a feasibility study, we trained CNNs to classify the area
28 surrounding bus stops (Lima, Peru) into *moderate* or *extreme* COVID-19 risk.

29 **Methods:** We used five images per bus stop. The outcome label (*moderate* or *extreme*) for each bus stop
30 was extracted from the local transport authority. We used transfer learning and updated the output layer of
31 five CNNs: NASNetLarge, InceptionResNetV2, Xception, ResNet152V2, and ResNet101V2. We chose the
32 best performing network which was further tuned. We used GradCam to understand the classification
33 process.

34 **Results:** NASNetLarge outperformed the other CNNs except in the recall metric for the *moderate* label and
35 in the precision metric for the *extreme* label; the ResNet152V2 performed better in these two metrics (85%
36 vs 76% and 63% vs 60%, respectively). The NASNetLarge was further tuned. The best recall (75%) and F1
37 score (65%) for the *extreme* label were reached with data augmentation techniques. Areas close to buildings
38 or with people were often classified as extreme risk.

39 **Conclusions:** This feasibility study showed that CNNs has the potential to classify street images according
40 to levels of COVID-19 risk. In addition to applications in clinical medicine, CNNs and street images could
41 advance the epidemiology of COVID-19 at the population level.

42
43 **Key words:** machine learning; deep learning; artificial intelligence; computer vision; built environment;
44 population health.

1
2
3 45 **Strengths and limitations of this study**
4

- 5
6 46 • We used five images per bus stop and the outcome information was provided by an official Government
7 institution.
8 47
9 48 • We leveraged on five well-known convolutional neural networks (transfer learning).
10 49
11 • The analysis focused on street images from one city only.
12 50
13 • Original data (street images) cannot be shared with the paper because of restricted access.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

51 INTRODUCTION

52 In COVID-19 research, deep learning tools applied to image analysis (i.e., computer vision) have informed
53 the diagnosis and prognosis of patients through classification of chest X-ray and computer tomography
54 images.¹⁻³ These tools have helped practitioners treating COVID-19 patients.

55 On the other hand, the application of computer vision to study the epidemiology of COVID-19 has been
56 limited. One relevant example is the use of Google Street View images to extract features of the built
57 environment and associate these with COVID-19 cases in the United States.⁴ This work showed that
58 unstructured and non-conventional data sources, such as street images, can deliver relevant information to
59 characterize the epidemiology of COVID-19 at the population level.⁴ However, computer vision models to
60 classify street images based on their COVID-19 risk do not exist. These models could be relevant to
61 understand unique local features of the built environment related to high COVID-19 risk. In addition, these
62 models could be applied to places where observed data are not available to identify whether this place is at
63 moderate or high risk of COVID-19, to inform relevant interventions. This could be particularly helpful in
64 Low- and Middle-income countries where limited resources do not allow massive COVID-19 testing, leaving
65 places with no observed information about the COVID-19 epidemiology.

66 In this pilot feasibility study, we aimed to ascertain whether a convolutional neural network (deep learning)
67 model could classify street images of bus stops according to their COVID-19 risk (binary outcome: *moderate*
68 versus *extreme* risk) in Lima, Peru. We also aimed to understand what features of the images were most
69 influential in the classification process. This is a pilot proof-of-concept empirical analysis.

70 METHODS

71 Study design

72 We used convolutional neural networks to study street images of bus stops and their surroundings in Lima,
73 Peru. We implemented a classification model to classify the bus stops into two labels: *moderate* or *extreme*
74 risk of COVID-19. We addressed a classification problem.

75 Rationale

76 We used five images per bus stop covering 360 degrees around the bus stop. Therefore, we targeted the
77 bus stop and the surrounding area. We did not target the bus stop itself alone. The bus stop was the anchor
78 for the outcome label (*moderate* or *extreme* risk of COVID-19) in the immediate surrounding area. It is
79 unlikely that COVID-19 risk would be confined to the bus stop itself. Rather, the bus stop would be a proxy
80 of the risk in the immediate nearby area.

81 We combined the five images before randomly splitting into the train, test and validation dataset. We used
82 the function *train_test_split* which randomly splits the data with equal distribution of the target outcome. We
83 did not condition the random split on the bus stops because we did not target the bus stop itself only. A
84 random split would provide data to have different profiles of the built environment, green areas, bus stops,
85 and other street features relevant for the model to learn and classify according to COVID-19 risk.

86 We deemed this a pilot feasibility proof-of-concept study because we aimed to provide preliminary data on
87 whether convolutional neural networks could classify street images according to COVID-19 risk. There is
88 abundant evidence about convolutional neural networks being used for classification of X-rays and other
89 clinical images. However, there is no evidence on convolutional neural networks being used for population
90 health. Ours is a proof-of-concept work, and future research could leverage on this idea with more images,
91 classifying into multiple outcome labels, and implementing more sophisticated networks.

92 Data sources

93 The labels (observed data) of the bus stops were downloaded from the website of the Authority for Urban
94 Transport in Lima and Callao (*Autoridad de Transporte Urbano para Lima y Callao* (ATU), name in Spanish).
95 This government office manages the public transportation service in Lima, and publishes a classification
96 map in which all bus stops in Lima are set into four categories of COVID-19 risk: moderate < high < very
97 high < extreme.⁵ Further details of their classification system have not been released. In this pilot feasibility
98 study, we only worked with the bus stops deemed as *moderate* (label 0) and *extreme* (label 1) risk of COVID-
99 19. We used the classification profile released on 2021-05-24.⁶ We conducted a pilot feasibility study
100 considering two outcome labels only, because we aimed to ascertain whether our hypothesis was possible
101 and lead to relevant results. Our feasibility proof-of-concept study could demonstrate that convolutional

1
2
3 102 neural networks could successfully classify street images according to COVID-19 risk. This has not been
4
5 103 studied before. Future work will leverage on this preliminary experience to develop a four-outcome model,
6
7 104 using larger datasets and incorporating more sophisticated networks.

8
9
10 105 We used the location (longitude and latitude coordinates) of the bus stops to download their street images
11
12 106 through the application programming interface (API) of Google Street View. That is, we downloaded all the
13
14 107 images in one batch through the API, rather than each one at the time through the API or from the standard
15
16 108 Google Street View website. For each bus stop (i.e., from each coordinate), we downloaded five images:
17
18 109 when the camera was facing at 0 degrees, at 90 degrees, at 180 degrees and at 270 degrees; in addition,
19
20 110 we also downloaded one image in which the direction of the camera was not specified (i.e., the heading
21
22 111 parameter in the API request was set at default). In other words, for each bus stop we had five images. We
23
24 112 did this to maximise the available data and to cover the surrounding area of the bus stop.⁷ Our rationale
25
26 113 was that the bus stop itself would not be responsible for the classification (moderate or extreme risk), but
27
28 114 the whole nearby environment.

115 **Original dataset**

116 Overall, after downloading both the labels and the images, there were 1,788 bus stops with their
117
118 corresponding label: 1,173 in the *moderate* category and 615 in the *extreme* category (1,173 + 615 = 1,788).
119
120 Because we used five images per bus stop, the analysis included 8,940 (1,788 x 5) images and their
121
122 corresponding label. The training dataset included a random sample of 60% (5,364) of the original dataset.
123
124 As further explained in the next section (Data preparation and class imbalance), after correcting for class
125
126 imbalance by introducing duplicates of the class with fewer observations, the training data included 7,024
127
128 observations (3,519 for *moderate* and 3,505 for *extreme* labels). The validation and test datasets included
129
130 a random sample of 20% of the original dataset each (0.20 x 8,940 = 1,788); the validation and test datasets
131
132 were not corrected for class imbalance.

133 **Data preparation and class imbalance**

134
135 126 We combined the images and the labels in one dataset, which was further divided into three datasets: the
136
137 127 training dataset including 60% of the data, the validation dataset including 20%, and the test dataset
138
139 128 including the remaining 20%. Data allocation to each of these three datasets was at random. After splitting

1
2
3 129 the data, we corrected for class imbalance in the train dataset only. We randomly multiplied the number of
4
5 130 images in the imbalanced outcome by 0.9. This led to virtually the same number of images for the *moderate*
6
7 131 and *extreme* risks labels.

8
9
10 132 There were two outcomes of interest: *moderate* and *extreme* risk. However, there were more observations
11
12 133 in the *moderate* category than in the *extreme* category. That is, there was class imbalance. After splitting
13
14 134 the data into the training, test and validation sets, we corrected for class imbalance in the training dataset
15
16 135 only. We randomly increased the number of observations in the extreme category by 90% in the training
17
18 136 dataset (not in validation and test datasets). The original (before correction for class imbalance) training set
19
20 137 had 3,519 observations in the *moderate* category and 1,845 in the *extreme* category ($3,519 + 1,845 = 5,346$).
21
22 138 After correcting for class imbalance as described before, the training dataset had 3,519 observations in the
23
24 139 *moderate* category (this number did not change) and 3,505 ($1.9 \times 1,845$) observations in the *extreme*
25
26 140 category. Therefore, there were 3,519 (*moderate*) + 3,505 (*extreme* after class imbalance correction) =
27
28 141 7,024 images and labels in total in the training dataset.

142 **Analysis**

143 In-depth details about the analysis are available in Supplementary Materials pp. 03-06. The analysis code
144 (Python Jupyter notebooks) is also available as Supplementary Materials.

145 In brief, in a pre-specified protocol we decided to elaborate on five deep convolutional neural networks pre-
146 trained with ImageNet (i.e., transfer learning). We chose these five networks because they have the best
147 top-5 accuracy of all models available in the Keras library:⁸ NASNetLarge, InceptionResNetV2, Xception,
148 ResNet152V2 and ResNet101V2. We implemented these five models with the same hyper-parameters, and
149 then we selected the one with the best performance which was further tuned and tested. The image
150 classification model was based on the latter model only (i.e., the one with the best performance out of the
151 five candidate models). We reported the loss and accuracy in the validation and test datasets; we also used
152 the test dataset to report the accuracy, recall and F1 score for each of the two possible outcomes (*moderate*
153 or *extreme* risk). Finally, we used GradCam to identify which areas of the input image were more relevant
154 to inform the classification process;⁹ for this, we randomly selected four images per outcome (i.e., four
155 images from the *moderate* label and four images from the *extreme* label).

156 Ethical statement

157 Human subjects were not directly studied in this work. The analysed data are in the public domain and can
158 be downloaded. The analysed data do not contain any personal identifiers. We did not seek approval by an
159 ethics committee.

160 Patient and public involvement

161 Human subjects did not participate nor were involved in this study.

162 RESULTS

163 Selection of the pre-trained model out of five candidate models

164 We used transfer learning and updated the output layer of five convolutional neural networks to predict our
165 two classes of interest. The NASNetLarge architecture and weights outperformed the other candidate
166 convolutional neural networks, except in the recall metric for the *moderate* label: 76% versus 85% in
167 NASNetLarge and ResNet152V2, respectively (Table 1). The ResNet152V2 also performed better than the
168 NASNetLarge in the precision metric for the *extreme* label (60% vs 63%). Further experiments were only
169 conducted with NASNetLarge because, overall, it performed better than the other pre-trained networks.

170 Model performance

171 We further tuned NASNetLarge with different hyper-parameters aiming to improve the accuracy (Table 2).
172 First, building on the initial hyper-parameters, we implemented two data augmentation options: horizontal
173 flip and zoom range. We chose these two data augmentation methods because they appropriately fit the
174 images under analysis; for example, because we were working with street images, a vertical flip would not
175 seem appropriate. The new model with horizontal flip improved the recall and F1 score for the *extreme* label;
176 from 68% with the original NASNetLarge to 75%, and from 64% to 65% (Figure 1). The new model with
177 horizontal flip and zoom range at 30% had better performance than the original NASNetLarge model in six
178 out of ten parameters, including precision for the *extreme* label.

179 Second, also building on the initial hyper-parameters (i.e., without data augmentation), the decay in the
180 stochastic gradient descent optimizer was changed from 1/25 (25 was the number of epochs) to 1/10

181 (the number of epochs was not changed). This model did not substantially improve the performance of the
182 model.

183 Third, building on the last specification (i.e., model with a decay of 1/10), we updated the monitoring factor
184 which updated the learning rate when it did not improve through epochs. Originally, this factor was 0.1, and
185 we updated it to 0.3. This model did not substantially improve the performance of the model.

186 **GradCam**

187 In the GradCam analysis we used the NASNetLarge model with one data augmentation technique
188 (horizontal flip). Even though the performance of the NASNetLarge model with two data augmentation
189 techniques (horizontal flip and zoom range) was better in more metrics, the model with horizontal flip only
190 had better recall and F1 score for the *extreme* label. The main indications for a *moderate* risk classification
191 were the presence of green areas and lack of close nearby buildings (Figure 2). Areas close to buildings or
192 with a considerable presence of people were often classified as extreme COVID-19 risk. The presence of
193 cars did not seem to impact the classification process.

194 **DISCUSSION**

195 **Main findings**

196 With almost all research on computer vision and COVID-19 focusing on diagnostic models based on X-rays
197 and other clinical images, our proof-of-concept work is novel because it borrows techniques from computer
198 vision into epidemiology and population health leveraging on available data (street images). This is the first
199 work of this kind, to the best of our knowledge. In this pilot feasibility proof-of-concept study, we showed
200 that deep convolutional neural networks can classify street images according to their COVID-19 risk with
201 acceptable accuracy. Future work should strengthen available convolutional neural networks or develop a
202 new architecture which could maximize the accuracy classification, not only for a binary outcome but also
203 covering multiple outcomes. This work could spark interest to use convolutional neural networks –and other
204 artificial intelligence tools– to advance population health and the epidemiological knowledge of COVID-19
205 (and other diseases), above and beyond the applications of convolutional neural networks for diagnosis and
206 prognosis of individual patients (e.g., classification of chest X-rays and compute tomography images¹⁻³).

207 Results in context

208 This feasibility proof-of-concept work signalled that a deep neural network is moderately accurate to classify
209 street images according to COVID-19 risk levels. These results are encouraging because the task we
210 pursued was difficult: to classify street images into levels for which there is no unique intrinsic information
211 in the images. Classification of, for example, chest X-ray images into healthy or ill could be easier for a
212 convolutional neural network because the X-ray of someone with a disease (e.g., pneumonia) would have
213 unique features (e.g., infiltrate spots at the bottom of the lungs) that a chest X-ray of someone healthy would
214 not have at all. Conversely, in our case, the street images did not have a unique underlying pattern to guide
215 the classification process. Our model had to work harder to find those unique characteristics to decide
216 between *moderate* and *extreme* risk.

217 Further tuning of the selected model (NASNetLarge) suggested that data augmentation methods improved
218 the performance of the model. When we updated the learning rate optimizer (decay and factor parameters),
219 the model performance did not substantially improve. This could suggest that for this particular task we may
220 need a large number of images. Alternatively, several combinations of data augmentation techniques would
221 need to be tested. Data augmentations should be carefully considered to select those most suitable for
222 these images; for example, vertical flip may not be a reasonable choice for street images.

223 Nguyen and colleagues used Google Street View images to associate features of the built environment with
224 COVID-19 cases in several states in the United States.⁴ Although we could have followed the same
225 approach, there would be some unique local features of the built environment that may not have been
226 identified by available object detection tools (e.g., street vendors and newspaper stands). We are not aware
227 of other peer-reviewed papers in which street images have been classified according to COVID-19
228 outcomes. Our work contributes to the available literature with a newly trained model benefiting from transfer
229 learning from a large and well-known architecture (NASNetLarge), based on images from a city in an upper-
230 middle income country (Lima, Peru).

231 The activation maps (GradCam) are not only useful to analyse the model's interpretation capability, but they
232 bolster the existing evidence of crowded places or indoor venues (such as nearby buildings) as COVID-19
233 high-risk areas; on the other hand, open spaces, such as green areas or locations far from buildings, remain

234 as moderate or low COVID-19 risk areas. Overall, our findings agree with the evidence describing crowded
235 areas, such as restaurants, gyms, hotels, and cafes, as having high COVID-19 transmission risk.¹⁰

236 **Public health implications**

237 Our work could have pragmatic applications to better understand the epidemiology of COVID-19 and to
238 inform public health measures. For example, our model –and future work improving this analysis– could be
239 used to characterize bus stops and other public places for which labelled data are not available. We worked
240 with images from bus stops in Lima, and our model could be applied to bus stops in other cities to
241 characterize their COVID-19 risk. Furthermore, our work could spark interest to conduct more sophisticated
242 analyses, like semantic segmentation whereby some unique elements of the local environment could be
243 identified as potential high-risk places. For example, bus stops often host food street vendors and
244 newspaper stands where people usually gather. Perhaps, the bus stops themselves are not high-risk places,
245 but the surrounding shops. This could inform policies and interventions to reduce the COVID-19 risk in these
246 places. Overall, deep learning techniques, including convolutional neural networks, could be adopted by
247 epidemiological research to advance the evidence about risk factors as well as disease outcomes and
248 distribution, in addition to their current use in clinical medicine.¹⁻³

249 As argued in the introduction, this is a pilot feasibility proof-of-concept study to study whether convolutional
250 neural networks could classify street images according to COVID-19 indicators. This work complements the
251 current use of convolutional neural networks for COVID-19 classification of clinical images (e.g., X-rays).
252 This work should be regarded as the first step in the use of convolutional neural networks in epidemiology
253 and population health relevant to COVID-19; this work is not the ultimate work on this subject and future
254 research should improve our approach and results.

255 **Ongoing and future work**

256 Ongoing and future work includes the development of a classification model for the four outcome labels
257 (i.e., *moderate*, *high*, *very high* and *extreme* COVID-19 risk). We will implement techniques that can
258 potentiate the classification capacity of the neural networks, including ensemble models,¹¹ novel loss
259 functions not currently implemented in the Keras environment (e.g., squared earth mover's distance-based
260 loss function),¹² and we may try other architectures (e.g., SqueezeNet¹³) with similar precision yet less

1
2
3 261 computationally expensive. Because most of our bus stop images also depicted buildings, we may try to
4
5 262 use a network already trained on images of buildings and other city landscapes (e.g., Places-365).
6
7

8 263 **Strengths and limitations**

9
10 264 In this preliminary proof-of-concept work, we followed a pre-defined protocol which included transfer learning
11
12 265 leveraging on large and deep neural networks trained with millions of images (ImageNet). We still had to
13
14 266 train the parameters of the output layer, for which we did not have a massive number of images. Future
15
16 267 work could expand our analysis with information and images from more bus stops or other public spaces to
17
18 268 train a more robust model. Ideally, these images should come from different cities. This information may be
19
20 269 available in other countries. There are further limitations we must acknowledge. First, the images and labels
21
22 270 were not synchronic; that is, the figures and the labels were not collected on the same date. This is a shared
23
24 271 limitation with other studies working with street images from open sources (e.g., Google Street View),
25
26 272 because these images are not taken continuously or in real time. This should not be a major limitation
27
28 273 because the analysis mostly focused on the built environment which has not changed substantially in recent
29
30 274 years. Because this feasibility study showed that the classification model performed moderately well,
31
32 275 researchers could collect new images in a prospective work to verify our findings with synchronic data. In
33
34 276 this line, satellite images collected daily could be useful. Second, we did not have exact details on how the
35
36 277 bus stops were classified by the local authorities. Nevertheless, we used official information which is
37
38 278 provided to the public for their safety and to inform them about the progression of the COVID-19 pandemic
39
40 279 in Lima, Peru. Therefore, we trust their method for classification is sound and based on the best available
41
42 280 evidence. Third, we had five images per bus stop: the fifth image did not look at a specific angle, unlike the
43
44 281 other four images that looked at 0, 90, 180 and 270 degrees around the bus stop. Therefore, the fifth image
45
46 282 had some overlap with the other images. We took this decision to maximize the available data. Researchers
47
48 283 with access to more labelled information, perhaps from public places overseas, could use the four images
49
50 284 without overlap and not significantly reducing the dataset size. In this line, the datasets (training, test and
51
52 285 validation) were split randomly and, just by chance albeit improbably, all images of one particular bus stop
53
54 286 could have fallen in a subset (e.g., test dataset). If so, the model would have poor accuracy to predict this
55
56 287 specific bus stop because the model did not have any information/images about that bus stop in the training
57
58 288 dataset. However, because we trained the model to classify *moderate* and *extreme* risk of COVID-19, the

1
2
3 289 model learnt patterns and profiles of the bus stops and their surrounding areas. This training could then be
4
5 290 applied to other bus stops with similar characteristics. The GradCam analysis helped us to exemplify the
6
7 291 patterns most influential in the selection process. Arguably, the influential patterns would be in all or most
8
9 292 images. Fourth, our model cannot be independently reproduced because we could not make the underlying
10
11 293 data available because these images do not belong to us. Google Street View images are available through
12
13 294 the API, though they need personal login credentials. Although this would not replace the raw underlying
14
15 295 data, to increase the transparency of our work we made available the Jupyter notebooks used in the analysis
16
17 296 (Supplementary Materials). These notebooks show the codes and results. Fifth, we did not report or discuss
18
19 297 the algorithms or computations behind the convolutional neural networks we used for transfer learning. As
20
21 298 per our protocol, we chose and applied a set of established convolutional neural networks to solve a
22
23 299 classification problem. Disentangling the underlying mechanisms underneath each convolutional neural
24
25 300 network was beyond the scope of this work. Nevertheless, it is relevant to understand the areas of the
26
27 301 images most influential in the classification process. This way we can verify if the classification process
28
29 302 followed a logical path. We therefore presented the GradCam analysis.

303 **Conclusions**

304 This pilot feasibility proof-of-concept study showed that a convolutional neural network has moderate
305 accuracy to classify street images into *moderate* and *extreme* risk of COVID-19. In addition to applications
306 in clinical medicine, deep convolutional neural networks have the potential to also advance the epidemiology
307 of COVID-19 at the population level exploring unstructured and non-conventional data sources.

308 TABLES

309

310 Table 1: Performance of the five candidate convolutional neural networks

311

	NASNetLarge	InceptionResNetV2	Xception	ResNet50	ResNet101V2
Loss, validation	0.526799	0.554040	0.533278	0.793278	0.744385
Accuracy, validation	0.742046	0.713636	0.730682	0.721027	0.723295
Loss, test	0.539906	0.557637	0.555917	0.800027	0.726274
Accuracy, test	0.731818	0.721591	0.706818	0.722227	0.718750
Precision, label 0 (<i>moderate</i>)	0.82	0.78	0.82	0.79	0.80
Recall, label 0 (<i>moderate</i>)	0.76	0.81	0.71	0.88	0.76
F1 score, label 0 (<i>moderate</i>)	0.79	0.79	0.76	0.88	0.78
Precision, label 1 (<i>extreme</i>)	0.60	0.61	0.56	0.63	0.58
Recall, label 1 (<i>extreme</i>)	0.68	0.56	0.70	0.48	0.64
F1 score, label 1 (<i>extreme</i>)	0.64	0.58	0.62	0.55	0.61

312

313 Green colour highlights the best metric, yellow colour highlights the second best metric and red colour highlights the third best metric row-wise. The
 314 precision, recall and F1 score are presented as proportions (multiply by 100 to have percentages). The precision, recall and F1 score were computed
 315 with the test dataset. Receiver Operating Characteristic (ROC) Curves for each model are available in Supplementary materials.

316 **Table 2: Further tuning of the selected model (NASNetLarge) and the performance metrics**

317

		New model specifications			
	Original model (as in Table1)	horizontal_flip = True // epochs = 25 (stopped at 12)	horizontal_flip = True // zoom_range = 0.30 // epochs = 25 (stopped at 15)	decay = 0.1/10 // epochs = 25 (stopped at 15)	decay = 0.1/10 // factor = 0.3 // epochs = 25 (stopped at 12)
Loss, validation	0.526799	0.534797	0.537553	0.532246	0.532246
Accuracy, validation	0.742046	0.737500	0.739773	0.732386	0.732386
Loss, test	0.539906	0.550286	0.528204	0.538252	0.538252
Accuracy, test	0.731818	0.719318	0.735795	0.725568	0.725568
Precision, label 0 (moderate)	0.82	0.85	0.76	0.83	0.83
Recall, label 0 (moderate)	0.76	0.71	0.87	0.74	0.74
F1 score, label 0 (moderate)	0.79	0.77	0.81	0.78	0.78
Precision, label 1 (extreme)	0.60	0.57	0.66	0.59	0.59
Recall, label 1 (extreme)	0.68	0.75	0.47	0.71	0.71
F1 score, label 1 (extreme)	0.64	0.65	0.55	0.64	0.64

318 Green colour highlights the best metric, yellow colour highlights the second best metric and red colour highlights the third best metric row-wise
 319 considering only the new model specifications. The precision, recall and F1 score are presented as proportions (multiply by 100 to have percentages).
 320 Receiver Operating Characteristic (ROC) Curves for each model are available in Supplementary Materials.

Downloaded from <http://bmjopen.bmj.com/> on June 10, 2025 at Agence Bibliographique de l'Enseignement Supérieur (ABES) . All rights reserved. No text and data mining, AI training, and similar technologies.

1
2
3
4
5 321 **FIGURES**

6
7
8 322 **Figure 1: Confusion matrix for the best NASNetLarge model**

9
10 323 This NASNetLarge model corresponds to the one with data augmentation of horizontal flip (first column in
11 324 the new model specification section of Table 2). The figure shows the absolute number of images in each
12 325 label: observed (true) on the y-axis and predicted on the x-axis.

13
14
15
16 326
17
18 327 **Figure 2: GradCam representation of four randomly selected images per outcome label (moderate**
19
20 328 **risk on the left and extreme risk on the right) and based on the final model**

21
22 329 Final model corresponds to the NASNetLarge showed in the first column in the new model specification
23 330 section of Table 2. The four images per outcome label were randomly chosen. From green to red, it shows
24 331 the areas that influenced the most in the classification process; that is, areas in red were the most
25
26 332 influential in the classification process. Images on the left column belong to the moderate label and
27
28 333 images on the right column belong to the extreme label. Images courtesy of Google Maps View.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 334 **CONTRIBUTIONS**

4
5 335 RMC-L and JFHSC conceived the idea. RMC-L conducted the analysis with support from JFHSC and MC-
6
7 336 C. MC-C supported the revised analysis. All authors approved the submitted version.
8
9
10 337

11
12
13 338 **FUNDING**

14 339 RMC-L is supported by a Wellcome Trust International Training Fellowship (Wellcome Trust
15
16 340 214185/Z/18/Z).
17
18
19 341

20
21
22
23 342 **COMPETING INTERESTS**

24 343 No conflict of interest.
25
26
27 344

28
29
30
31 345 **DATA AVAILABILITY STATEMENT**

32 346 Outcome (i.e., labels: moderate and extreme COVID-19 risk) data are available online:
33
34 347 <https://sistemas.atu.gob.pe/paraderosCOVID>; this information was systematized at
35
36 348 <https://github.com/jmcastagnetto/lima-atu-covid19-paraderos>. The images were downloaded from Google
37
38 349 Street View through the API with a personal account; images cannot be shared with third parties. All analysis
39
40 350 codes are available as Python Jupyter Notebooks in Supplementary Materials. JupyterLab Notebooks and
41
42 351 the final model (weights) are available at:
43
44 352 [https://figshare.com/articles/online_resource/Street_images_classification_according_to_COVID-](https://figshare.com/articles/online_resource/Street_images_classification_according_to_COVID-19_risk_in_Lima_Peru_A_convolutional_neural_networks_feasibility_analysis/17321021)
45
46 353 [19_risk_in_Lima_Peru_A_convolutional_neural_networks_feasibility_analysis/17321021](https://figshare.com/articles/online_resource/Street_images_classification_according_to_COVID-19_risk_in_Lima_Peru_A_convolutional_neural_networks_feasibility_analysis/17321021)
47
48
49 354

355 **REFERENCES**

- 356 1. Ghaderzadeh M, Asadi F. Deep Learning in the Detection and Diagnosis of COVID-19 Using
357 Radiology Modalities: A Systematic Review. *Journal of healthcare engineering* 2021; **2021**: 6677314.
- 358 2. Mohammad-Rahimi H, Nadimi M, Ghalyanchi-Langeroudi A, Taheri M, Ghafouri-Fard S.
359 Application of Machine Learning in Diagnosis of COVID-19 Through X-Ray and CT Images: A Scoping
360 Review. *Frontiers in cardiovascular medicine* 2021; **8**: 638011.
- 361 3. Montazeri M, ZahediNasab R, Farahani A, Mohseni H, Ghasemian F. Machine Learning Models
362 for Image-Based Diagnosis and Prognosis of COVID-19: Systematic Review. 2021; **9**(4): e25181.
- 363 4. Nguyen QC, Huang Y, Kumar A, et al. Using 164 Million Google Street View Images to Derive
364 Built Environment Predictors of COVID-19 Cases. 2020; **17**(17).
- 365 5. Autoridad de Transporte Urbano para Lima y Callao (ATU). Paraderos con Riesgo de COVID -
366 19. URL: <https://sistemas.atu.gob.pe/paraderosCOVID>.
- 367 6. URL: <https://github.com/jmcastagnetto/lima-atu-covid19-paraderos>.
- 368 7. Suel E, Polak JW, Bennett JE, Ezzati M. Measuring social, environmental and health inequalities
369 using deep learning and street imagery. *Scientific reports* 2019; **9**(1): 6229.
- 370 8. Keras applications. URL: <https://keras.io/api/applications/>.
- 371 9. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual
372 Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer
373 Vision* 2020; **128**(2): 336-59.
- 374 10. Chang S, Pierson E, Koh PW, et al. Mobility network models of COVID-19 explain inequities and
375 inform reopening. *Nature* 2021; **589**(7840): 82-7.
- 376 11. Ganaie M, Hu M. Ensemble deep learning: A review. *arXiv preprint arXiv:210402395* 2021.
- 377 12. Hou L, Yu C-P, Samaras D. Squared earth mover's distance-based loss for training deep neural
378 networks. *arXiv preprint arXiv:161105916* 2016.
- 379 13. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level
380 accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:160207360* 2016.

381

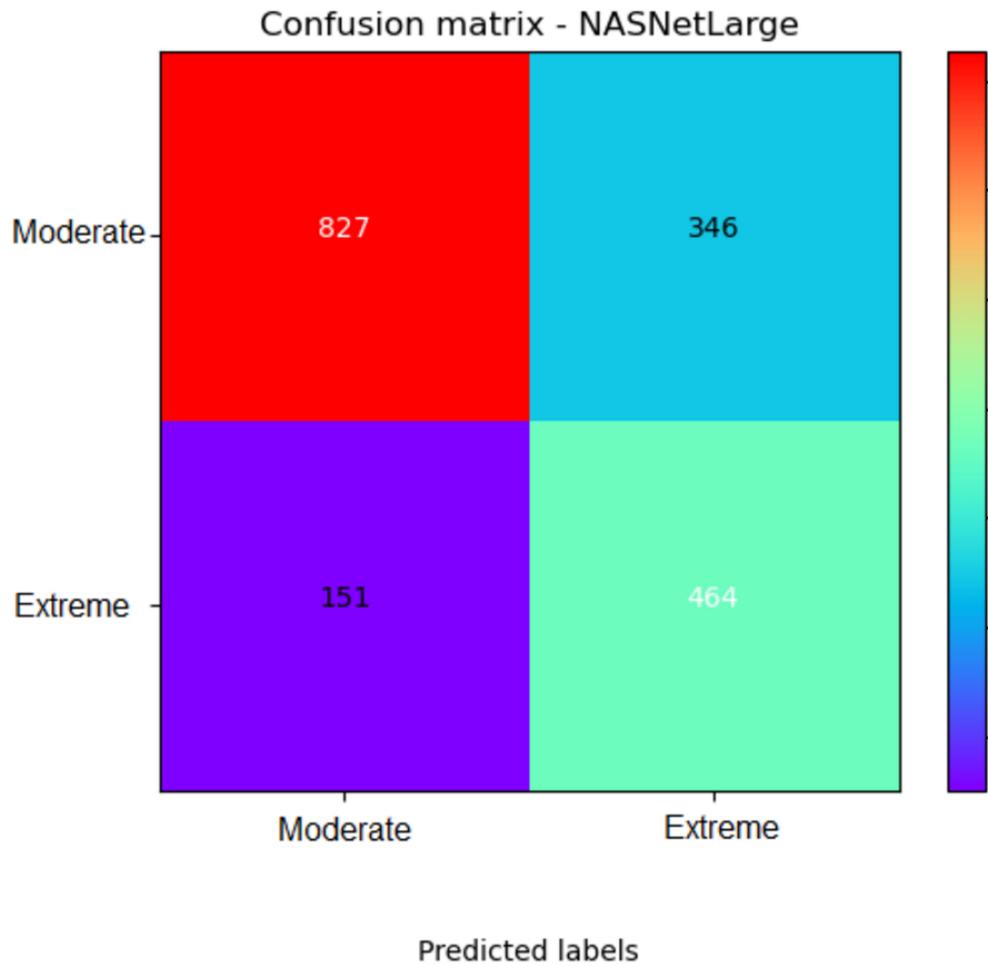


Figure 1: Confusion matrix for the best NASNetLarge model

This NASNetLarge model corresponds to the one with data augmentation of horizontal flip (first column in the new model specification section of Table 2). The figure shows the absolute number of images in each label: observed (true) on the y-axis and predicted on the x-axis.

601x601mm (38 x 38 DPI)

Supplementary Materials

Street images classification according to COVID-19 risk in Lima, Peru: A convolutional neural networks feasibility analysis

Corresponding author:

Rodrigo M Carrillo-Larco, MD

Department of Epidemiology and Biostatistics

School of Public Health

Imperial College London

rcarrill@ic.ac.uk

Contents

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

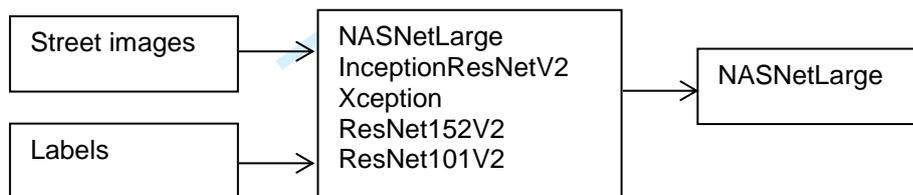
EXPANDED METHODS	3
Overview	3
System details	3
Images.....	3
Labels (outcome)	4
Class imbalance	4
Image data generator.....	4
Convolutional neural networks.....	5
Model Architecture	5
Model Training	6
Activation Heatmap by GradCam Visualization	7
References.....	7
Receiver Operating Characteristic (ROC) Curves I	9
Receiver Operating Characteristic (ROC) Curves II	10

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Supérieur (ABES).

EXPANDED METHODS

Overview

In a pre-specified protocol we decided to test five well-known convolutional neural network architectures. These five networks are those with the best accuracy among the models available in the Keras library.¹ From these five candidate models, we chose the one with the best performance for our task; this model was further tuned to improve its performance.



Analysis code (Jupyter notebooks) and the weights of the final model are found here:

<https://drive.google.com/file/d/1HXLsenn7yvxri7n2xE80WMtxQzi5fgBX/view?usp=sharing>

System details

Analyses were conducted with a GPU NVIDIA Quadro P1000 on Python Jupyter (version 3.7.10). The notebooks are provided as supplementary materials.

Images

The list of bus stop in Lima, Peru, and their COVID-19 risk, are provided by local transport authority.² This information has been extracted and homogenized and is available online.³ Key for our work, this information contains: i) location of each bus stop (latitude and longitude coordinates), which was used to extract street images; and ii) the COVID-19 risk level assigned to each bus stop, which was used as the outcome labels.

We downloaded the images from Google Street View through the application programming interface (API). We used the Python libraries *google_streetview.api* and *request*. For each bus stop (i.e., for each latitude and longitude coordinate), the API request specified the heading parameter = [0, 90, 180, 270]; in addition, we downloaded one image for which the heading parameter was set at default. Consequently, for each bus stop we had five images in total. The images were downloaded with size 640x640 pixels. In this feasibility study there were 1,788 bus stops, and because we had five images per bus stop, we used 8,940 (1,788 x 5) images in total.

Labels (outcome)

We used the bus stop classification released on 2021-05-24,^{2, 3} by the local transport authority in Lima, Peru.² They classify the bus stops in four categories of COVID-19 risk: moderate, high, very high and extreme risk.² Details on how they made this classification are not available. Nevertheless, because this is official information released by a public authority to inform the general population, we trust their classification is based on the best available evidence. In this feasibility work we only used bus stops labelled as *moderate* (n=1,173) and *extreme* (n=615) COVID-19 risk. There were 1,788 (1,173 + 615) bus stops in total. The list of labels was appended five times, so that we would have as many images as labels (NB: we had five images per bus stop as described above).

Class imbalance

There were two outcomes of interest: moderate and extreme risk. However, there were more observations in the moderate category than in the extreme category. That is, there was class imbalance. After splitting the data into the training, test and validation sets, we corrected for class imbalance. We randomly increased the number of observations in the extreme category by 90% in the training dataset (not in validation and test datasets). The original (before correction for class imbalance) training set had 3,519 observations in the moderate category and 1,845 in the extreme category (3,519 + 1,845 = 5,364). After correcting for class imbalance as described before, the training dataset had 3,519 observations in the moderate category (this number did not change) and 3,505 (~1.9 x 1,845) observations in the extreme category (3,519 + 3,505 = 7,024 total sample in the training dataset).

Image data generator

We constructed a dataframe with two columns: the path to each image (i.e., to the exact location where the images were saved) and the corresponding label for each image. This dataframe was passed to the *ImageDataGenerator* function of the *keras.preprocessing.function* library. At this point, we also re-scaled the images between 0 and 1 by dividing by 255. Then, we created three image iterators: one for the training, validation and test datasets (*train_datagen.flow_from_dataframe* function). To the *train_datagen.flow_from_dataframe* function we passed the dataframe with the location of the images and their labels, specified this was a categorical classification problem, a batch size of 32, and a

specific image size for each candidate model (see table below); in addition, the image iterator for the training dataset had the shuffle parameter as *True*.

Convolutional neural networks

We decided to train five candidate convolutional neural networks with the following specifications.

	NASNetLarge	InceptionResNetV2	Xception	ResNet152V2	ResNet101V2
Image size	331 x 331	299 x 299		224 x 224	
Pre-trained weights	ImageNet				
Top layer included?	No				
Trainable parameters	None				
Additional layers	Dense layer with 2 neurons (for the two outcome labels), with <i>softmax</i> activation				
Number of epochs for training	25				
Optimizer	SGD(learning_rate = 0.1, momentum = 0.9, decay = 0.1/number of epochs, nesterov = True)				
Loss function	Binary crossentropy				

SDG: stochastic descent gradient.

In addition, we monitored the validation loss: when the validation loss would not improve in one epoch, then the learning rate was multiplied by 0.1. We also specified an early stop: when the validation loss would not improve for ten epochs, the training would stop. To choose between the five candidate models we did not implement any data augmentation methods.

We chose the model with the best performance (Table 1 in the main text), which was further tuned (Table 2 in main text).

Model Architecture

The chosen model (NASNetLarge) presented in this article is a state-of-the-art convolutional neural network (CNN) pre-trained with the ImageNet dataset of images. The NASNetLarge neural network is widely used in computer vision. It is composed of several layers of convolutional cells that extract the most important features from each image, learning which features are the most characteristic from each image category. Most pre-trained state-of-the-art CNNs, such as the AlexNet or the ResNet50, base their innovations in the use of complex activation functions, efficient designs and special layers, such as the Batch Normalization,⁴ which not only improve accuracy performance, but also reduce the time needed to train a model. The ResNet50 neural network, for example, introduces the concept of residual neural networks, layers that skip their connections to other layers by adding connection shortcuts, thus reducing backpropagation training time and better generalizing models.

1
2
3 However, most of these CNNs had complex design that was built by using trial-and-error methods,
4 oblivious to modern state-of-the-art optimization methods such as the use of evolutionary and nature-
5 inspired algorithms or Reinforcement Learning (RL). This is where Neural Search Architecture (NAS)
6 Networks come in play.⁵ Dubbed as a breakthrough in machine learning automation, NAS networks
7 use *AI for AI*. The network's whole architecture is designed by optimization algorithms, such as
8 Gradient-based search and RL. The idea behind this is that, setting the right restrictions to avoid
9 repetitive inclusion of layers, the network cannot only be trained by its parameters, but by its own
10 architecture, adding and changing layers, activations functions and connections.
11
12
13
14
15
16
17
18

19 Our research uses the NASNetLarge model available by Keras,¹ pre-trained on the ImageNet dataset
20 on 1000 categories. Because we only have two categories to train on (moderate and extreme), we
21 started by replacing the network's fully connected classification layer by a 2-neurons layer, with a
22 softmax activation function.¹ The rest of the original NASNetLarge network was kept unmodified to
23 preserve its proven architecture.
24
25
26
27
28
29

30 **Model Training**

31 To train our model we used transfer learning. Based on the pre-trained NASNet model, we retained its
32 parameters and froze them, keeping just trainable the last fully connected layer, initializing its
33 parameters randomly by He uniform initialization.⁶ This single-stage training took advantage of the
34 pre-trained parameters speeding up training by just focusing on the last decision layer. The model
35 was trained for 30 epochs. Each epoch is understood as a complete training cycle through the whole
36 train dataset. Data is then fed by batches; once all batches are loaded, an epoch is finished. By using
37 a batch size of 32, training and testing sets are fed to the neural network. The loss function, as we are
38 dealing with a binary classification model, is binary cross entropy.
39
40
41
42
43
44
45
46
47

48 As optimizer we used the Stochastic Gradient Descent (SGD).⁷ One of the key advantages of this
49 optimizer is due to its stochasticity in selecting each batch for the backpropagation step. Although the
50 model takes longer to converge, unlike other optimizers, the SGD has been proven to converge better
51 local optima, searching for global optima with much more easiness. This factor helps also to reduce
52 overfitting.
53
54
55
56
57
58
59
60

Activation Heatmap by GradCam Visualization

Model interpretability is a feature that has gained importance since the appearance of methods such as GradCam.⁸ This technique uses the values from the gradients in the model's final feature layer to produce visual explanations, highlighting regions of importance taken by the model to infer a given input. In other words, this technique informs which areas of the input figure were more relevant to make the final classification.

Regions that represent higher gradient values, accounting for most of the last layer activations and network's decision, are represented by being coloured closer to the red portion of the spectrum. By comparison, regions with the lowest activations, not adding much information to the network's final decision, appear as areas closer to the blue portion of the spectrum. These gradient values are taken from the network's last convolutional layer, right before the last pooling layer.

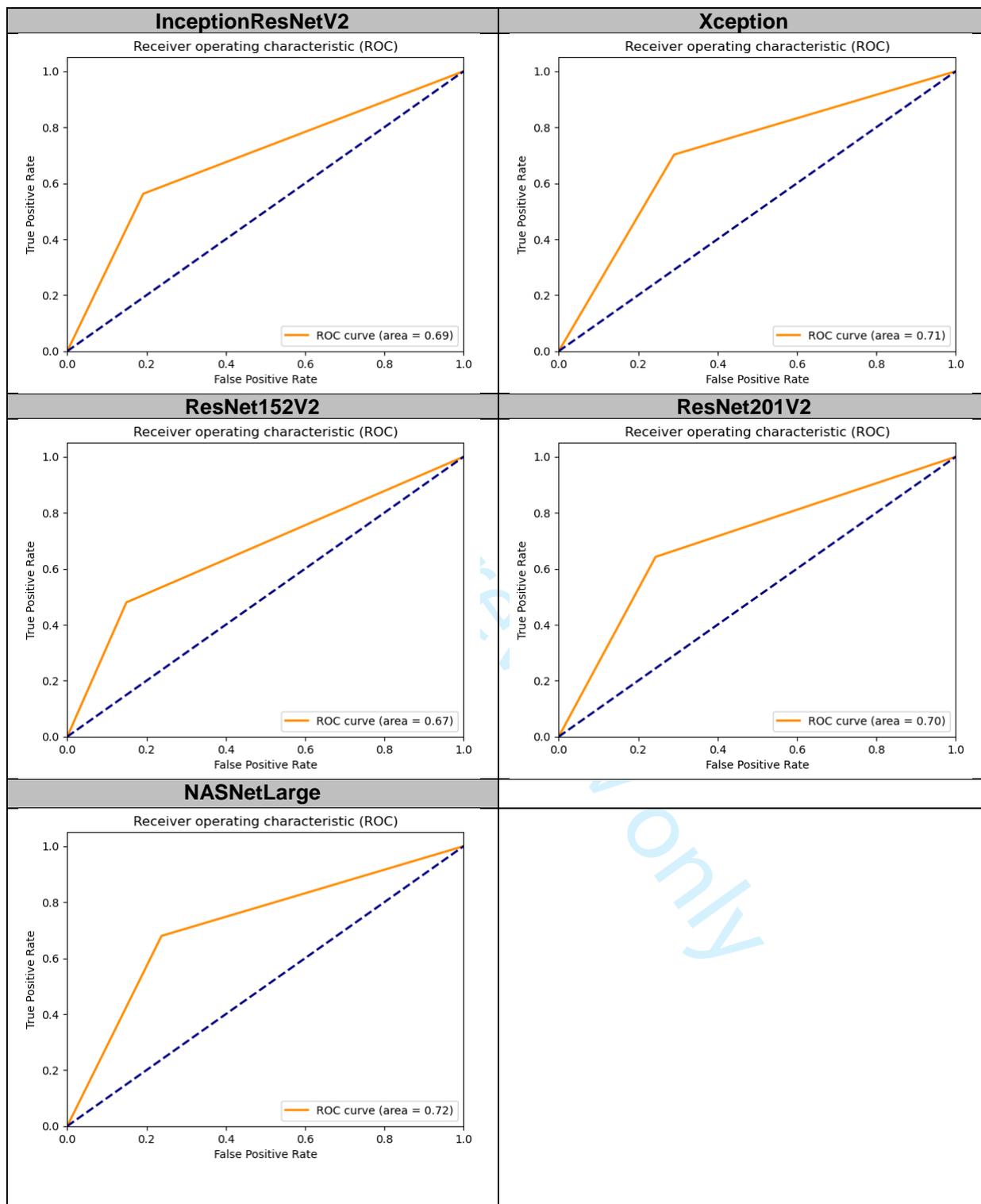
References

1. Keras applications. URL: <https://keras.io/api/applications/>.
2. Autoridad de Transporte Urbano para Lima y Callao (ATU). Paraderos con Riesgo de COVID - 19. URL: <https://sistemas.atu.gob.pe/paraderosCOVID>.
3. URL: <https://github.com/jmcastagnetto/lima-atu-covid19-paraderos>.
4. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv e-prints* 2015: arXiv:1502.03167.
5. Kyriakides G, Margaritis K. An Introduction to Neural Architecture Search for Convolutional Networks. *arXiv e-prints* 2020: arXiv:2005.11074.
6. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *arXiv e-prints* 2015: arXiv:1502.01852.
7. Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv e-prints* 2016: arXiv:1608.03983.
8. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision* 2020; **128**(2): 336-59.

For peer review only

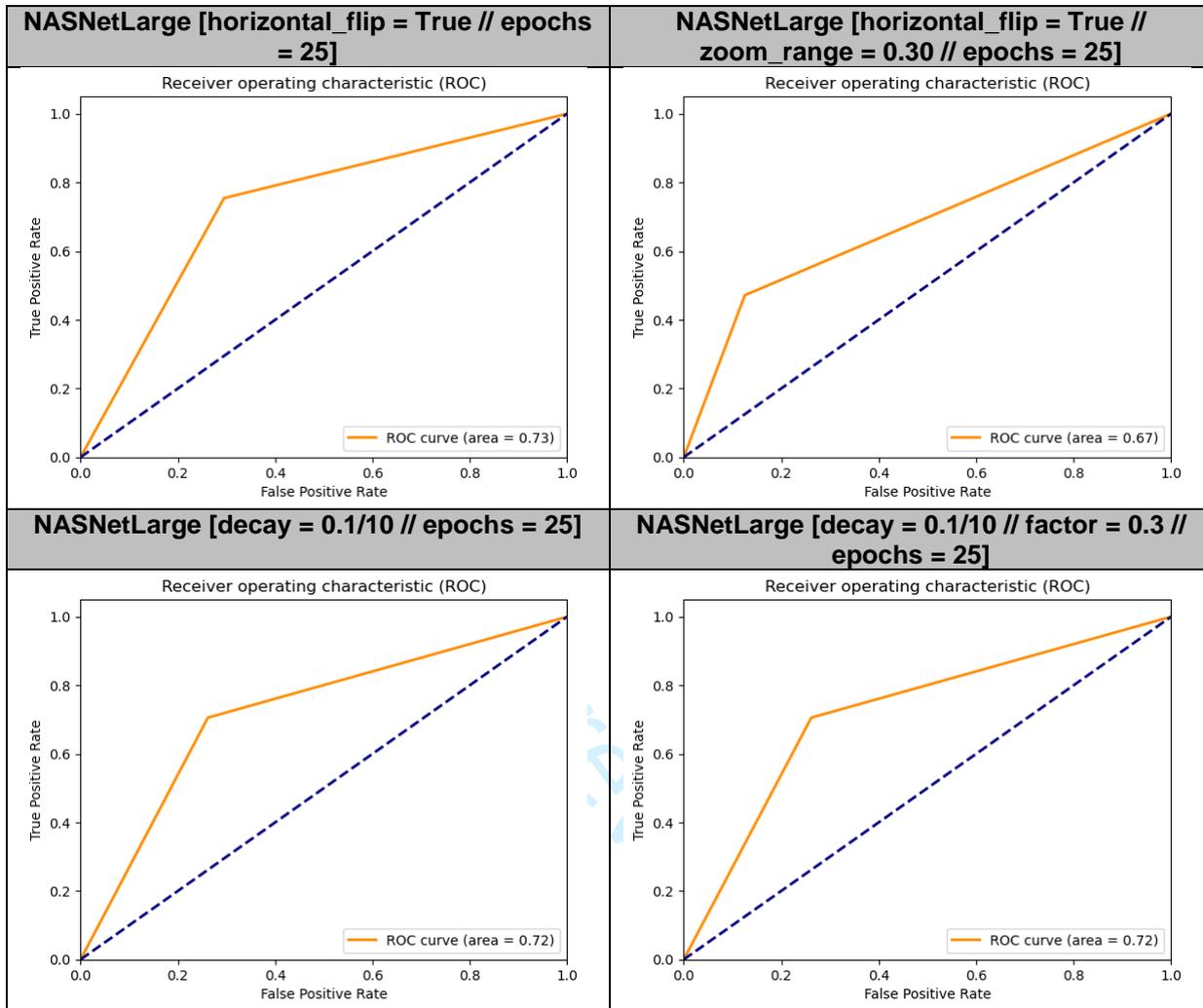
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Receiver Operating Characteristic (ROC) Curves I



Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

Receiver Operating Characteristic (ROC) Curves II



Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

only

BMJ Open

Street images classification according to COVID-19 risk in Lima, Peru: A convolutional neural networks feasibility analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-063411.R1
Article Type:	Original research
Date Submitted by the Author:	27-Jul-2022
Complete List of Authors:	Carrillo-Larco, Rodrigo; Imperial College London, Castillo-Cara, Manuel; Universidad Politécnica de Madrid, Ontology Engineering Group Hernández Santa Cruz, Jose Francisco; Independent Researcher
Primary Subject Heading:	Global health
Secondary Subject Heading:	Epidemiology
Keywords:	COVID-19, EPIDEMIOLOGY, PUBLIC HEALTH

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3
4
5 1 **Street images classification according to COVID-19 risk in Lima, Peru: A convolutional neural**
6
7 2 **networks feasibility analysis**
8
9 3

10
11 4 Rodrigo M Carrillo-Larco^{1,2,3}
12
13

14 5 Manuel Castillo-Cara^{4,5}
15
16

17 6 Jose Francisco Hernández Santa Cruz⁶
18
19

- 20 7
21
22 8 1. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London,
23 London, UK
24 9
25 10 2. CRONICAS Centre of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia,
26 Lima, Peru
27 11
28 12 3. Universidad Continental, Lima, Peru
29 13
30 14 4. Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
31 15
32 16 5. Universidad de Lima, Lima, Peru
33 17
34 18 6. Independent researcher, Edinburgh, UK
35 19
36 20
37 21
38 22

39 16
40
41 17 **Corresponding author:**
42
43

44 18 Rodrigo M Carrillo-Larco, MD
45

46 19 Department of Epidemiology and Biostatistics
47

48 20 School of Public Health
49

50 21 Imperial College London
51

52 22 rcarrill@ic.ac.uk
53
54
55
56
57
58
59
60

23 ABSTRACT

24 **Objectives:** During the COVID-19 pandemic, convolutional neural networks (CNNs) have been used in
25 clinical medicine (e.g., chest X-rays classification). Whether CNNs could inform the epidemiology of COVID-
26 19 classifying street images according to COVID-19 risk is unknown, yet it could pinpoint high-risk places
27 and relevant features of the built environment. In a feasibility study, we trained CNNs to classify the area
28 surrounding bus stops (Lima, Peru) into *moderate* or *extreme* COVID-19 risk.

29 **Design:** CNN analysis based on images from bus stops and the surrounding area. We used transfer learning
30 and updated the output layer of five CNNs: NASNetLarge, InceptionResNetV2, Xception, ResNet152V2,
31 and ResNet101V2. We chose the best performing CNN which was further tuned. We used GradCam to
32 understand the classification process.

33 **Setting:** Bus stops from Lima, Peru. We used five images per bus stop.

34 **Primary and secondary outcome measures:** Bus stop images were classified according to COVID-19 risk
35 into two labels: *moderate* or *extreme*.

36 **Results:** NASNetLarge outperformed the other CNNs except in the recall metric for the *moderate* label and
37 in the precision metric for the *extreme* label; the ResNet152V2 performed better in these two metrics (85%
38 vs 76% and 63% vs 60%, respectively). The NASNetLarge was further tuned. The best recall (75%) and F1
39 score (65%) for the *extreme* label were reached with data augmentation techniques. Areas close to buildings
40 or with people were often classified as extreme risk.

41 **Conclusions:** This feasibility study showed that CNNs have the potential to classify street images according
42 to levels of COVID-19 risk. In addition to applications in clinical medicine, CNNs and street images could
43 advance the epidemiology of COVID-19 at the population level.

44
45 **Key words:** machine learning; deep learning; artificial intelligence; computer vision; built environment;
46 population health.

1
2
3 **47 Strengths and limitations of this study**
4

- 5
6 48 • We used five images per bus stop and the outcome information was provided by an official Government
7
8 49 institution.
9
10 50 • We leveraged on five well-known convolutional neural networks (transfer learning).
11
12 51 • The analysis focused on street images from one city only.
13
14 52 • Original data (street images) cannot be shared because of restricted access.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

53 INTRODUCTION

54 In COVID-19 research, deep learning tools applied to image analysis (i.e., computer vision) have informed
55 the diagnosis and prognosis of patients through classification of chest X-ray and computer tomography
56 images.¹⁻³ These tools have helped practitioners treating COVID-19 patients.

57 On the other hand, the application of computer vision to study the epidemiology of COVID-19 has been
58 limited. One relevant example is the use of Google Street View images to extract features of the built
59 environment and associate these with COVID-19 cases in the United States of America.⁴ This work showed
60 that unstructured and non-conventional data sources, such as street images, can deliver relevant
61 information to characterize the epidemiology of COVID-19 at the population level.⁴ In a similar vein, though
62 not exclusively addressing COVID-19, other researchers have leveraged on street images to study health-
63 related social inequalities,⁵ air pollution,⁶ walkability,⁷ as well as the built environment and health outcomes.^{8,9}
64 These examples show the potential of computer vision for population health research, above and beyond
65 its multiple applications in clinical medicine with diagnostic and prognostic models.

66 However, to the best of our knowledge, computer vision models to classify street images based on their
67 COVID-19 risk do not exist. From a public health perspective, such models could be relevant to understand
68 unique local features of the built environment related to high COVID-19 risk. In addition, these models could
69 be applied to places where observed data are not available to identify whether this place is at moderate or
70 high risk of COVID-19 and inform potential interventions. This could be particularly helpful in Low- and
71 Middle-income countries where limited resources do not allow massive COVID-19 testing, leaving places
72 with no observed information about the COVID-19 epidemiology, though the local epidemiology could be
73 estimated based on available images or alternative sources.

74 In this pilot feasibility study, we aimed to ascertain whether a convolutional neural network (deep learning)
75 model could classify street images of bus stops according to their COVID-19 risk (binary outcome: *moderate*
76 versus *extreme* risk) in Lima, Peru. We also aimed to understand what features of the images were most
77 influential in the classification process.

78 METHODS

79 Study design

80 We used convolutional neural networks to study street images of bus stops and their surroundings in Lima,
81 Peru. We implemented a classification model to classify the bus stops into two labels: *moderate* or *extreme*
82 risk of COVID-19. We addressed a classification problem.

83 Rationale

84 We used five images per bus stop covering 360 degrees around the bus stop. Therefore, we targeted the
85 bus stop and the surrounding area. We did not target the bus stop itself alone. The bus stop was the anchor
86 for the outcome label (*moderate* or *extreme* risk of COVID-19) in the immediate surrounding area. It is
87 unlikely that COVID-19 risk would be confined to the bus stop itself. Rather, the bus stop would be a proxy
88 of the risk in the immediate nearby area.

89 We combined the five images before randomly splitting into the train, test, and validation dataset. We used
90 the function *train_test_split* which randomly splits the data with equal distribution of the target outcome. We
91 did not condition the random split on the bus stops because we did not target the bus stop itself only. A
92 random split would provide data to have different profiles of the built environment, green areas, bus stops,
93 and other street features relevant for the model to learn and classify according to COVID-19 risk.

94 We deemed this a pilot feasibility proof-of-concept study because we aimed to provide preliminary data on
95 whether convolutional neural networks could classify street images according to COVID-19 risk. While there
96 is evidence about convolutional neural networks being used for classification of X-rays and other clinical
97 images for COVID-19 diagnosis,¹⁻³ there is less evidence on convolutional neural networks being used for
98 population health and COVID-19. Future research could leverage on this idea with more images, classifying
99 into multiple outcome labels, and implementing more sophisticated networks.

100 Public health and epidemiological research usually rely on structured data sources such as health surveys
101 and measurements from patients including samples such as blood. Unstructured data sources, such as
102 images, are gaining attention in clinical medicine and have been used to develop diagnosis and prognostic
103 models; however, the use of images, including street images, in public health and epidemiological research

1
2
3 104 is limited. This work elaborates on this premise and on the current burden by COVID-19 and was conceived
4
5 105 to study whether street images can ascertain the COVID-19 risk in the community. If successful, a deep
6
7 106 learning model to classify street images according to COVID-19 risk could be used for disease surveillance,
8
9 107 and to estimate the risk in places where observed data lack.

108 **Data sources**

109 The labels (observed data) of the bus stops were downloaded from the website of the Authority for Urban
110 Transport in Lima and Callao (*Autoridad de Transporte Urbano para Lima y Callao* (ATU), name in Spanish).
111 This government office manages the public transportation service in Lima, and publishes a classification
112 map in which all bus stops in Lima are set into four categories of COVID-19 risk: moderate < high < very
113 high < extreme.¹⁰ Although this is an official source of information from a government branch, details of how
114 the bus stops were classified are not available; please, refer to the discussion section where we further
115 elaborate on this caveat. In this pilot feasibility study, we only worked with the bus stops deemed as
116 *moderate* (label 0) and *extreme* (label 1) risk of COVID-19. We used the classification profile released on
117 2021-05-24.¹¹ We conducted a pilot feasibility study considering two outcome labels only. This, because we
118 aimed to ascertain whether our hypothesis was possible and lead to relevant results while studying, from a
119 public health perspective, the most important outcomes signalling the extremes of the risk distribution.
120 Developing a model to identify areas at moderate risk could signal places where restrictions can be relaxed
121 or suspended. Similarly, developing a model to identify areas at extreme risk would signal places where
122 restrictions should be kept or strengthened. Therefore, a model focusing on two labels only, where these
123 labels represent the extremes of the risk distribution, would be relevant and provide actional evidence. Our
124 study could demonstrate that convolutional neural networks could successfully classify street images
125 according to COVID-19 risk, with not addition information such as number of cases or health determinants.
126 This has not been studied before. Future work will leverage on this preliminary experience to develop a four-
127 outcome model, using larger datasets and incorporating more sophisticated networks.

128 We used the location (longitude and latitude coordinates) of the bus stops to download their street images
129 through the application programming interface (API) of Google Street View. That is, we downloaded all the
130 images in one batch through the API, rather than each one at the time through the API or from the standard

1
2
3 131 Google Street View website. For each bus stop (i.e., from each coordinate), we downloaded five images:
4
5 132 when the camera was facing at 0 degrees, at 90 degrees, at 180 degrees and at 270 degrees; in addition,
6
7 133 we also downloaded one image in which the direction of the camera was not specified (i.e., the heading
8
9 134 parameter in the API request was set at default). In other words, for each bus stop we had five images. We
10
11 135 did this to maximise the available data and to cover the surrounding area of the bus stop.¹² Our rationale
12
13 136 was that the bus stop itself would not be responsible for the classification (moderate or extreme risk), but
14
15 137 the whole nearby environment. Consequently, if the bus stop was labelled as moderate or extreme risk, the
16
17 138 same label applied to the images of the surrounding area. For example, if bus stop X was labelled as
18
19 139 moderate risk, all five images for such bus stop were labelled as moderate risk (i.e., image of the bus stop
20
21 140 itself plus the four images of the surrounding area).

141 **Original dataset**

142 Overall, after downloading both the labels and the images, there were 1,788 bus stops with their
143 corresponding label: 1,173 in the *moderate* category and 615 in the *extreme* category ($1,173 + 615 = 1,788$).
144 Because we used five images per bus stop, the analysis included 8,940 ($1,788 \times 5$) images and their
145 corresponding label. The training dataset included a random sample of 60% (5,364) of the original dataset.
146 As further explained in the next section (Data preparation and class imbalance), after correcting for class
147 imbalance by introducing duplicates of the class with fewer observations, the training data included 7,024
148 observations (3,519 for *moderate* and 3,505 for *extreme* labels). The validation and test datasets included
149 a random sample of 20% of the original dataset each ($0.20 \times 8,940 = 1,788$); the validation and test datasets
150 were not corrected for class imbalance.

151 **Data preparation and class imbalance**

152 We combined the images and the labels in one dataset, which was further divided into three datasets: the
153 training dataset including 60% of the data, the validation dataset including 20%, and the test dataset
154 including the remaining 20%. Data allocation to each of these three datasets was at random. After splitting
155 the data, we corrected for class imbalance in the train dataset only. We randomly multiplied the number of
156 images in the imbalanced outcome by 0.9. This led to virtually the same number of images for the *moderate*
157 and *extreme* risks labels.

1
2
3 158 There were two outcomes of interest: *moderate* and *extreme* risk. However, there were more observations
4
5 159 in the *moderate* category than in the *extreme* category. That is, there was class imbalance. After splitting
6
7 160 the data into the training, test, and validation sets, we corrected for class imbalance in the training dataset
8
9 161 only. We randomly increased the number of observations in the extreme category by 90% in the training
10
11 162 dataset (not in validation and test datasets). The original (before correction for class imbalance) training set
12
13 163 had 3,519 observations in the *moderate* category and 1,845 in the *extreme* category (3,519 + 1,845 = 5,346).
14
15 164 After correcting for class imbalance as described before, the training dataset had 3,519 observations in the
16
17 165 *moderate* category (this number did not change) and 3,505 (1.9 x 1,845) observations in the *extreme*
18
19 166 category. Therefore, there were 3,519 (*moderate*) + 3,505 (*extreme* after class imbalance correction) =
20
21 167 7,024 images and labels in total in the training dataset.
22

23 168 **Analysis**

24
25 169 In-depth details about the analysis are available in Supplementary Materials pp. 03-06. The analysis code
26
27 170 (Python Jupyter notebooks) is also available as Supplementary Materials.

28
29
30 171 In brief, in a pre-specified protocol we decided to elaborate on five deep convolutional neural networks pre-
31
32 172 trained with ImageNet (i.e., transfer learning). We chose these five networks because they have the best
33
34 173 top-5 accuracy of all models available in the Keras library:¹² NASNetLarge, InceptionResNetV2, Xception,
35
36 174 ResNet152V2 and ResNet101V2. We implemented these five models with the same hyper-parameters, and
37
38 175 then we selected the one with the best performance which was further tuned and tested. The image
39
40 176 classification model was based on the latter model only (i.e., the one with the best performance out of the
41
42 177 five candidate models). We reported the loss and accuracy in the validation and test datasets; we also used
43
44 178 the test dataset to report the accuracy, recall and F1 score for each of the two possible outcomes (*moderate*
45
46 179 or *extreme* risk). Finally, we used GradCam (class activation maps) to identify which areas of the input image
47
48 180 were more relevant to inform the classification process;¹³ for this, we randomly selected four images per
49
50 181 outcome (i.e., four images from the *moderate* label and four images from the *extreme* label). Areas most
51
52 182 activated as shown by brighter colours, would be decisively in the classification process.
53

54 183 **Patient and public involvement**

55
56 184 Human subjects did not participate nor were involved in this study.
57
58
59
60

185 RESULTS

186 Selection of the pre-trained model out of five candidate models

187 We used transfer learning and updated the output layer of five convolutional neural networks to predict our
188 two classes of interest. The NASNetLarge architecture and weights outperformed the other candidate
189 convolutional neural networks, except in the recall metric for the *moderate* label: 76% versus 85% in
190 NASNetLarge and ResNet152V2, respectively (Table 1). The ResNet152V2 also performed better than the
191 NASNetLarge in the precision metric for the *extreme* label (60% vs 63%). Further experiments were only
192 conducted with NASNetLarge because, overall, it performed better than the other pre-trained networks.

193

194 **Table 1: Performance of the five candidate convolutional neural networks**

195

	NASNetLarge	InceptionResNetV2	Xception	ResNet152V2	ResNet101V2
Loss, validation	0.526799	0.554040	0.533278	0.793147	0.744385
Accuracy, validation	0.742046	0.713636	0.730682	0.721023	0.723295
Loss, test	0.539906	0.557637	0.555917	0.800661	0.726274
Accuracy, test	0.731818	0.721591	0.706818	0.722727	0.718750
Precision, label 0 (moderate)	0.82	0.78	0.82	0.76	0.80
Recall, label 0 (moderate)	0.76	0.81	0.71	0.85	0.76
F1 score, label 0 (moderate)	0.79	0.79	0.76	0.80	0.78
Precision, label 1 (extreme)	0.60	0.61	0.56	0.63	0.58
Recall, label 1 (extreme)	0.68	0.56	0.70	0.48	0.64
F1 score, label 1 (extreme)	0.64	0.58	0.62	0.54	0.61

196

197 Green colour highlights the best metric, yellow colour highlights the second best metric and red colour
 198 highlights the third best metric row-wise. The precision, recall and F1 score are presented as proportions
 199 (multiply by 100 to have percentages). The precision, recall and F1 score were computed with the test
 200 dataset. Receiver Operating Characteristic (ROC) Curves for each model are available in Supplementary
 201 Materials.

1
2
3 202 **Model performance**

4
5 203 We further tuned NASNetLarge with different hyper-parameters aiming to improve the accuracy (Table 2).
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

204 **Table 2: Further tuning of the selected model (NASNetLarge) and the performance metrics**

205

		New model specifications			
	Original model (as in Table1)	horizontal_flip = True // epochs = 25 (stopped at 12)	horizontal_flip = True // zoom_range = 0.30 // epochs = 25 (stopped at 15)	decay = 0.1/10 // epochs = 25 (stopped at 12)	decay = 0.1/10 // factor = 0.3 // epochs = 25 (stopped at 12)
Loss, validation	0.526799	0.534797	0.537553	0.532246	0.532246
Accuracy, validation	0.742046	0.737500	0.739773	0.732386	0.732386
Loss, test	0.539906	0.550286	0.528204	0.538252	0.538252
Accuracy, test	0.731818	0.719318	0.735795	0.725568	0.725568
Precision, label 0 (moderate)	0.82	0.85	0.76	0.83	0.83
Recall, label 0 (moderate)	0.76	0.71	0.87	0.74	0.74
F1 score, label 0 (moderate)	0.79	0.77	0.81	0.78	0.78
Precision, label 1 (extreme)	0.60	0.57	0.66	0.59	0.59
Recall, label 1 (extreme)	0.68	0.75	0.47	0.71	0.71
F1 score, label 1 (extreme)	0.64	0.65	0.55	0.64	0.64

206 Green colour highlights the best metric, yellow colour highlights the second best metric and red colour
 207 highlights the third best metric row-wise considering only the new model specifications. The precision,
 208 recall and F1 score are presented as proportions (multiply by 100 to have percentages). Receiver
 209 Operating Characteristic (ROC) Curves for each model are available in Supplementary Materials.

1
2
3 210 First, building on the initial hyper-parameters, we implemented two data augmentation options: horizontal
4
5 211 flip and zoom range. We chose these two data augmentation methods because they appropriately fit the
6
7 212 images under analysis; for example, because we were working with street images, a vertical flip would not
8
9 213 seem appropriate. The new model with horizontal flip improved the recall and F1 score for the *extreme* label;
10
11 214 from 68% with the original NASNetLarge to 75%, and from 64% to 65% (Figure 1). The new model with
12
13 215 horizontal flip and zoom range at 30% had better performance than the original NASNetLarge model in six
14
15 216 out of ten parameters, including precision for the *extreme* label.

16
17 217 Second, also building on the initial hyper-parameters (i.e., without data augmentation), the decay in the
18
19 218 stochastic gradient descent optimizer was changed from 1/25 (25 was the number of epochs) to 1/10
20
21 219 (the number of epochs was not changed). This model did not substantially improve the performance of the
22
23 220 model.

24
25
26 221 Third, building on the last specification (i.e., model with a decay of 1/10), we updated the monitoring factor
27
28 222 which updated the learning rate when it did not improve through epochs. Originally, this factor was 0.1, and
29
30 223 we updated it to 0.3. This model did not substantially improve the performance of the model.

31 32 224 **GradCam**

33
34 225 In the GradCam (i.e., class activation maps) analysis we used the NASNetLarge model with one data
35
36 226 augmentation technique (horizontal flip). Even though the performance of the NASNetLarge model with two
37
38 227 data augmentation techniques (horizontal flip and zoom range) was better in more metrics, the model with
39
40 228 horizontal flip only had better recall and F1 score for the *extreme* label. The main indications for a *moderate*
41
42 229 risk classification were the presence of green areas and lack of close nearby buildings. That is, images with
43
44 230 several open spaces like parks, open streets, or wide avenues, would most likely be classified as moderate
45
46 231 risk. Conversely, areas close to buildings, with a considerable presence of people, and with meeting points
47
48 232 (e.g., street vendors), were often classified as extreme COVID-19 risk. In other words, bus stops with one
49
50 233 or multiple street vendors, newspapers stand, or any other point for people to gather around, would most
51
52 234 likely be classified as extreme risk. The presence of cars did not seem to impact the classification process.

235 DISCUSSION

236 Main findings

237 With almost all research on computer vision and COVID-19 focusing on diagnostic models based on X-rays
238 and other clinical images, our work is novel because it borrows techniques from computer vision into
239 epidemiology and population health leveraging on available data (street images). In this study we showed
240 that deep convolutional neural networks can classify street images according to their COVID-19 risk with
241 acceptable accuracy. Future work should strengthen available convolutional neural networks or develop a
242 new architecture which could maximize the accuracy classification, not only for a binary outcome but also
243 covering multiple outcomes. This work could spark interest to use convolutional neural networks –and other
244 artificial intelligence tools– to advance population health and the epidemiological knowledge of COVID-19
245 (and other diseases), above and beyond the applications of convolutional neural networks for diagnosis and
246 prognosis of individual patients (e.g., classification of chest X-rays and compute tomography images¹⁻³).

247 Results in context

248 This work signalled that a deep neural network is moderately accurate to classify street images according
249 to COVID-19 risk levels. These results are encouraging because the task we pursued was difficult: to classify
250 street images into levels for which there is no unique intrinsic information in the images. Classification of,
251 for example, chest X-ray images into healthy or ill could be easier for a convolutional neural network because
252 the X-ray of someone with a disease (e.g., pneumonia) would have unique features (e.g., infiltrate spots at
253 the bottom of the lungs) that a chest X-ray of someone healthy would not have at all. Conversely, in our
254 case, the street images did not have a unique underlying pattern to guide the classification process. Our
255 model had to work harder to find those unique characteristics to decide between *moderate* and *extreme*
256 risk.

257 Further tuning of the selected model (NASNetLarge) suggested that data augmentation methods improved
258 the performance of the model. When we updated the learning rate optimizer (decay and factor parameters),
259 the model performance did not substantially improve. This could suggest that, for this particular task, we
260 may need a large number of images. Alternatively, several combinations of data augmentation techniques

261 would need to be tested. Data augmentations should be carefully considered to select those most suitable
262 for these images; for example, vertical flip may not be a reasonable choice for street images.

263 Nguyen and colleagues used Google Street View images to associate features of the built environment with
264 COVID-19 cases in several states in the United States of America.⁴ Although we could have followed the
265 same approach, there would be some unique local features of the built environment that may not have been
266 identified by available object detection tools (e.g., street vendors and newspaper stands). We are not aware
267 of other peer-reviewed papers in which street images have been classified according to COVID-19
268 outcomes developing a new model or leveraging on transfer knowledge from an established neural network.
269 Our work contributes to the available literature with a newly trained model benefiting from transfer learning
270 from a large and well-known architecture (NASNetLarge), based on images from a city in an upper-middle
271 income country (Lima, Peru).

272 The activation maps (GradCam results) are not only useful to analyse the model's interpretation capability,
273 but they bolster the existing evidence of crowded places or indoor venues (such as nearby buildings) as
274 COVID-19 high-risk areas. For example, areas with street vendors would activate more than open spaces
275 for the extreme risk classification; on the other hand, open areas would play a major role in classifying
276 moderate risk images. Overall, our findings agree with the evidence describing crowded areas, such as
277 restaurants, gyms, hotels, and cafes, as having high COVID-19 transmission risk.¹⁴ Furthermore, our work
278 advances the field by showing that street images with no other clinical or epidemiological data, have
279 moderate accuracy to predict COVID-19 risk.

280 **Public health implications**

281 Our work could have pragmatic applications to better understand the epidemiology of COVID-19 and to
282 inform public health interventions. For example, our model –and future work improving this analysis– could
283 be used to characterize bus stops and other public places for which labelled data are not available. We
284 worked with images from bus stops in Lima, and our model could be applied to bus stops in other cities to
285 characterize their COVID-19 risk, particularly where observed data are not available. Furthermore, our work
286 could spark interest to conduct more sophisticated analyses, like semantic segmentation whereby some
287 unique elements of the local environment could be identified as potential high-risk places. For example, bus

1
2
3 288 stops in Lima often host food street vendors and newspaper stands where people usually gather. Perhaps,
4
5 289 the bus stops themselves are not high-risk places, but those surrounding shops. This could inform policies
6
7 290 and interventions to reduce the COVID-19 risk in these places. Overall, deep learning techniques, including
8
9 291 convolutional neural networks, could be adopted by epidemiological research to advance the evidence about
10
11 292 risk factors as well as disease outcomes and distribution, in addition to their current use in clinical medicine.¹⁻

12
13 293 ³

14
15
16 294 Our work was designed to understand whether and how well street images, without complementary data,
17
18 295 can predict COVID-19 risk. Our results support the idea that the built environment alone is a health
19
20 296 determinant because the street images were not complemented with other epidemiological data such as
21
22 297 number of cases or COVID-19 transmission. Measuring COVID-19 throughout a country can be challenging
23
24 298 and barriers include lack of access to tests as well as laboratory facilities to process the samples, and limited
25
26 299 health or trained personnel to take the samples. Our work suggests that street images could serve as proxy
27
28 300 to estimate the COVID-19 risk in places where this information does not exist based on observed data.
29
30 301 Therefore, we provide preliminary evidence suggesting that street images can be instrumental in COVID-19
31
32 302 surveillance.

33
34 303 Finally, as argued before, this is a pilot feasibility proof-of-concept study to study whether convolutional
35
36 304 neural networks could classify street images according to COVID-19 indicators. This work complements the
37
38 305 current use of convolutional neural networks for COVID-19 classification of clinical images (e.g., X-rays).
39
40 306 This work should be regarded as the first step in the use of convolutional neural networks in epidemiology
41
42 307 and population health relevant to COVID-19; this work is not the ultimate work on this subject and future
43
44 308 research should improve our approach and results.

45 46 309 **Ongoing and future work**

47
48 310 Ongoing and future work includes the development of a classification model for the four outcome labels (i.e.,
49
50 311 *moderate, high, very high* and *extreme* COVID-19 risk). We will implement techniques that can potentiate
51
52 312 the classification capacity of the neural networks, including ensemble models,¹⁵ novel loss functions not
53
54 313 currently implemented in the Keras environment (e.g., squared earth mover's distance-based loss

1
2
3 314 function),¹⁶ and we may try other architectures (e.g., SqueezeNet¹⁷) with similar precision yet less
4
5 315 computationally expensive. Because most of our bus stop images also depicted buildings, we may try to
6
7 316 use a network already trained on images of buildings and other city landscapes (e.g., Places-365).
8
9

10 317 **Strengths and limitations**

11 318 We followed a pre-defined protocol which included transfer learning leveraging on large and deep neural
12
13 319 networks trained with millions of images (ImageNet). We still had to train the parameters of the output layer,
14
15 320 for which we did not have a massive number of images. Future work could expand our analysis with
16
17 321 information and images from more bus stops or other public spaces to train a more robust model. Ideally,
18
19 322 these images should come from different cities. This information may be available in other countries. There
20
21 323 are further limitations we must acknowledge. First, the images and labels were not synchronic; that is, the
22
23 324 figures and the labels were not collected on the same date. This is a shared limitation with other studies
24
25 325 working with street images from open sources (e.g., Google Street View), because these images are not
26
27 326 taken continuously or in real time. This should not be a major limitation because the analysis mostly focused
28
29 327 on the built environment which has not changed substantially in recent years. Because this feasibility study
30
31 328 showed that the classification model performed moderately well, researchers could collect new images in a
32
33 329 prospective work to verify our findings with synchronic data. In this line, satellite images collected daily could
34
35 330 be useful. Second, we did not have exact details on how the bus stops were classified by the local
36
37 331 authorities. Nevertheless, we used official information which is provided to the public for their safety and to
38
39 332 inform them about the progression of the COVID-19 pandemic. Because it is an official source of public
40
41 333 information, we trust their method for classification is sound and based on the best available evidence. This
42
43 334 limitation should not substantially bias our model or results because the labels were clearly available from
44
45 335 the data provider (transport authority), and we did not have to make any assumptions nor manual labelling.
46
47 336 However, this may limit the external reproducibility of our work because other researchers may not label
48
49 337 their images following the same criteria by our data source. We argue that this should not rest importance
50
51 338 to our work because which could serve as basis for future research in the area in which the underlying
52
53 339 labelling criteria are clearer. Third, we had five images per bus stop: the fifth image did not look at a specific
54
55 340 angle, unlike the other four images that looked at 0, 90, 180 and 270 degrees around the bus stop.
56
57 341 Therefore, the fifth image had some overlap with the other images. We took this decision to maximize the
58
59
60

1
2
3 342 available data. Researchers with access to more labelled information, perhaps from public places overseas,
4
5 343 could use the four images without overlap and not significantly reducing the dataset size. In this line, the
6
7 344 datasets (training, test, and validation) were split randomly and, just by chance albeit improbably, all images
8
9 345 of one particular bus stop could have fallen in a subset (e.g., test dataset). If so, the model would have poor
10
11 346 accuracy to predict this specific bus stop because the model did not have any information/images about that
12
13 347 bus stop in the training dataset. However, because we trained the model to classify *moderate* and *extreme*
14
15 348 risk of COVID-19, the model learnt patterns and profiles of the bus stops and their surrounding areas. This
16
17 349 training could then be applied to other bus stops with similar characteristics. The GradCam analysis helped
18
19 350 us to exemplify the patterns most influential in the selection process. Arguably, the influential patterns would
20
21 351 be in all or most images. Fourth, our model cannot be independently reproduced because we could not
22
23 352 make the underlying data available because these images do not belong to us. Google Street View images
24
25 353 are available through the API, though they need personal login credentials. Although this would not replace
26
27 354 the raw underlying data, to increase the transparency of our work we made available the Jupyter notebooks
28
29 355 used in the analysis (Supplementary Materials). These notebooks show the codes and results. Fifth, we did
30
31 356 not report or discuss the algorithms or computations behind the convolutional neural networks we used for
32
33 357 transfer learning. As per our protocol, we chose and applied a set of established convolutional neural
34
35 358 networks to solve a classification problem. Disentangling the underlying mechanisms underneath each
36
37 359 convolutional neural network was beyond the scope of this work. Nevertheless, it is relevant to understand
38
39 360 the areas of the images most influential in the classification process. This way we can verify if the
40
41 361 classification process followed a logical path. We therefore reported the GradCam analysis.

42 362 **Conclusions**

43
44 363 This study showed that a convolutional neural network has moderate accuracy to classify street images into
45
46 364 *moderate* and *extreme* risk of COVID-19. In addition to applications in clinical medicine, deep convolutional
47
48 365 neural networks have the potential to also advance the epidemiology of COVID-19 at the population level
49
50 366 exploding unstructured and non-conventional data sources.

1
2
3
4 367 **FIGURES**

5
6
7 368 **Figure 1: Confusion matrix for the best NASNetLarge model**

8
9 369 This NASNetLarge model corresponds to the one with data augmentation of horizontal flip (first column in
10
11 370 the new model specification section of Table 2). The figure shows the absolute number of images in each
12
13 371 label: observed (true) on the y-axis and predicted on the x-axis.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 372 **CONTRIBUTIONS**

4
5 373 RMC-L and JFHSC conceived the idea. RMC-L conducted the analysis with support from JFHSC and MC-
6
7 374 C. MC-C supported the revised analysis. All authors approved the submitted version.
8
9

10 375

11
12
13 376 **FUNDING**

14 377 RMC-L is supported by a Wellcome Trust International Training Fellowship (Wellcome Trust
15
16 378 214185/Z/18/Z).
17
18

19 379

20
21
22
23 380 **COMPETING INTERESTS**

24 381 No conflict of interest.
25
26

27 382

28
29
30
31 383 **DATA AVAILABILITY STATEMENT**

32 384 Outcome (i.e., labels: moderate and extreme COVID-19 risk) data are available online:
33
34 385 <https://sistemas.atu.gob.pe/paraderosCOVID>; this information was systematized at
35
36 386 <https://github.com/jmcastagnetto/lima-atu-covid19-paraderos>. The images were downloaded from Google
37
38 387 Street View through the API with a personal account; images cannot be shared with third parties. All analysis
39
40 388 codes are available as Python Jupyter Notebooks in Supplementary Materials. JupyterLab Notebooks and
41
42 389 the final model (weights) are available at:

43
44
45 390 [https://figshare.com/articles/online_resource/Street_images_classification_according_to_COVID-
46
47 391 19_risk_in_Lima_Peru_A_convolutional_neural_networks_feasibility_analysis/17321021](https://figshare.com/articles/online_resource/Street_images_classification_according_to_COVID-19_risk_in_Lima_Peru_A_convolutional_neural_networks_feasibility_analysis/17321021)
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 392 **Ethical statement**

4
5 393 Human subjects were not directly studied in this work. The analysed data are in the public domain and can
6
7 394 be downloaded. The analysed data do not contain any personal identifiers. We did not seek approval by an
8
9 395 ethics committee.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

396 **REFERENCES**

- 397 1. Ghaderzadeh M, Asadi F. Deep Learning in the Detection and Diagnosis of COVID-19 Using
398 Radiology Modalities: A Systematic Review. *Journal of healthcare engineering* 2021; **2021**: 6677314.
- 399 2. Mohammad-Rahimi H, Nadimi M, Ghalyanchi-Langeroudi A, Taheri M, Ghafouri-Fard S. Application
400 of Machine Learning in Diagnosis of COVID-19 Through X-Ray and CT Images: A Scoping Review.
401 *Frontiers in cardiovascular medicine* 2021; **8**: 638011.
- 402 3. Montazeri M, ZahediNasab R, Farahani A, Mohseni H, Ghasemian F. Machine Learning Models for
403 Image-Based Diagnosis and Prognosis of COVID-19: Systematic Review. 2021; **9**(4): e25181.
- 404 4. Nguyen QC, Huang Y, Kumar A, et al. Using 164 Million Google Street View Images to Derive Built
405 Environment Predictors of COVID-19 Cases. 2020; **17**(17).
- 406 5. Suel E, Polak JW, Bennett JE, Ezzati M. Measuring social, environmental and health inequalities
407 using deep learning and street imagery. *Scientific reports* 2019; **9**(1): 6229.
- 408 6. Suel E, Sorek-Hamer M, Moise I, et al. What You See Is What You Breathe? Estimating Air Pollution
409 Spatial Variation Using Street-Level Imagery. *Remote Sensing* 2022; **14**(14): 3429.
- 410 7. Nagata S, Nakaya T, Hanibuchi T, Amagasa S, Kikuchi H, Inoue S. Objective scoring of streetscape
411 walkability related to leisure walking: Statistical modeling approach with semantic segmentation of Google
412 Street View images. *Health & place* 2020; **66**: 102428.
- 413 8. Nguyen QC, Keralis JM, Dwivedi P, et al. Leveraging 31 Million Google Street View Images to
414 Characterize Built Environments and Examine County Health Outcomes. *Public Health Rep* 2021; **136**(2):
415 201-11.
- 416 9. Nguyen QC, Sajjadi M, McCullough M, et al. Neighbourhood looking glass: 360° automated
417 characterisation of the built environment for neighbourhood effects research. *Journal of epidemiology and
418 community health* 2018; **72**(3): 260-6.
- 419 10. Autoridad de Transporte Urbano para Lima y Callao (ATU). Paraderos con Riesgo de COVID - 19.
420 July 22, 2022. <https://sistemas.atu.gob.pe/paraderosCOVID>.
- 421 [dataset]11. jmcstaghetto. Data from: lima-atu-covid19-paraderos. July 22, 2022.
422 <https://github.com/jmcstaghetto/lima-atu-covid19-paraderos>.
- 423 12. Keras applications. July 26, 2022. <https://keras.io/api/applications/>.

- 1
2
3 424 13. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations
4
5 425 from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision* 2020;
6
7 426 **128**(2): 336-59.
- 8
9 427 14. Chang S, Pierson E, Koh PW, et al. Mobility network models of COVID-19 explain inequities and
10
11 428 inform reopening. *Nature* 2021; **589**(7840): 82-7.
- 12
13 429 15. Ganaie M, Hu M. Ensemble deep learning: A review. *arXiv preprint arXiv:210402395* 2021.
- 14
15 430 16. Hou L, Yu C-P, Samaras D. Squared earth mover's distance-based loss for training deep neural
16
17 431 networks. *arXiv preprint arXiv:161105916* 2016.
- 18
19 432 17. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level
20
21 433 accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:160207360* 2016.

434

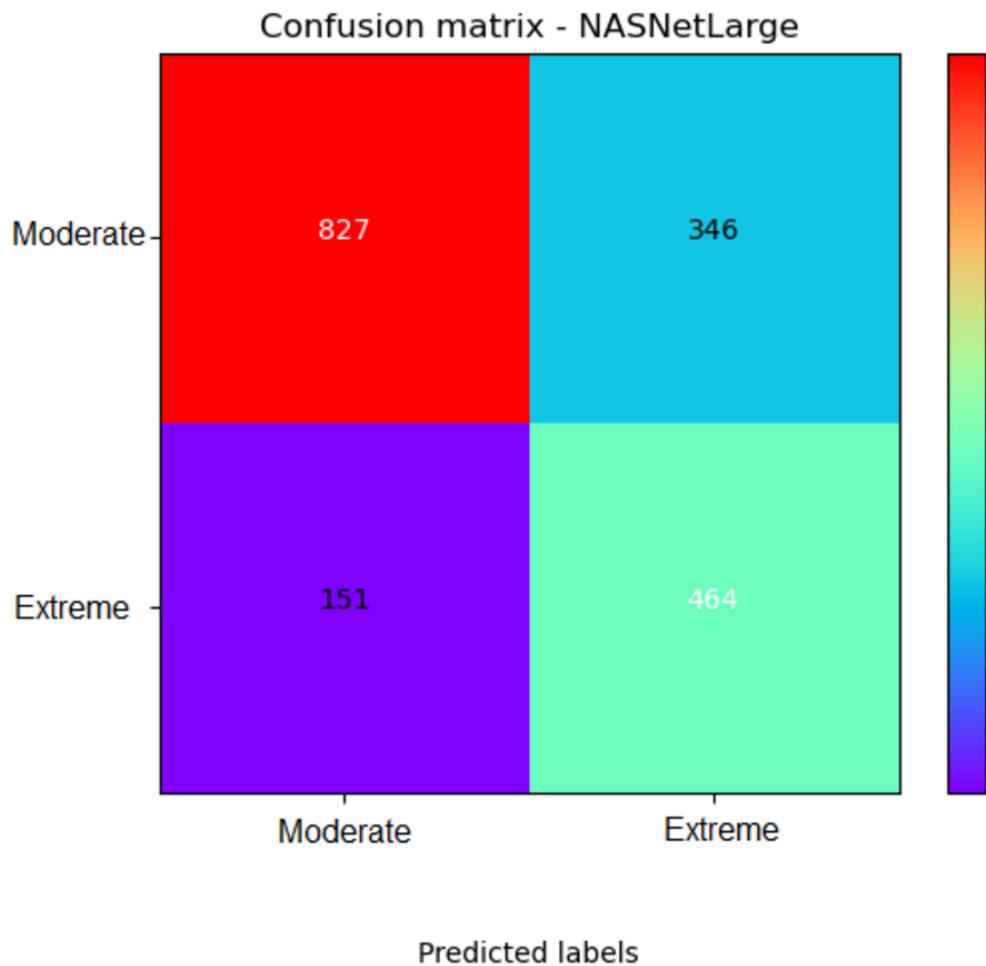


Figure 1: Confusion matrix for the best NASNetLarge model

This NASNetLarge model corresponds to the one with data augmentation of horizontal flip (first column in the new model specification section of Table 2). The figure shows the absolute number of images in each label: observed (true) on the y-axis and predicted on the x-axis.

601x601mm (38 x 38 DPI)

Supplementary Materials

Street images classification according to COVID-19 risk in Lima, Peru: A convolutional neural networks feasibility analysis

Corresponding author:

Rodrigo M Carrillo-Larco, MD

Department of Epidemiology and Biostatistics

School of Public Health

Imperial College London

rcarrill@ic.ac.uk

Contents

1

2

3

4

5 EXPANDED METHODS 3

6

7 Overview 3

8

9 System details 3

10

11 Images..... 3

12

13 Labels (outcome) 4

14

15 Class imbalance 4

16

17 Image data generator 4

18

19 Convolutional neural networks 5

20

21 Model Architecture 5

22

23 Model Training 6

24

25 Activation Heatmap by GradCam Visualization 7

26

27 References 7

28

29

30

31

32

33 Receiver Operating Characteristic (ROC) Curves I 9

34

35

36 Receiver Operating Characteristic (ROC) Curves II 10

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

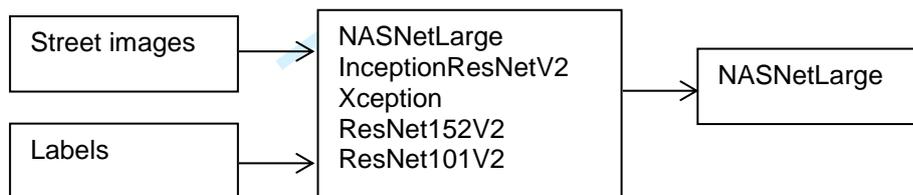
60

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

EXPANDED METHODS

Overview

In a pre-specified protocol we decided to test five well-known convolutional neural network architectures. These five networks are those with the best accuracy among the models available in the Keras library.¹ From these five candidate models, we chose the one with the best performance for our task; this model was further tuned to improve its performance.



Analysis code (Jupyter notebooks) and the weights of the final model are found here:

<https://drive.google.com/file/d/1HXLsenn7yvxri7n2xE80WMtxQzi5fgBX/view?usp=sharing>

System details

Analyses were conducted with a GPU NVIDIA Quadro P1000 on Python Jupyter (version 3.7.10). The notebooks are provided as supplementary materials.

Images

The list of bus stop in Lima, Peru, and their COVID-19 risk, are provided by local transport authority.² This information has been extracted and homogenized and is available online.³ Key for our work, this information contains: i) location of each bus stop (latitude and longitude coordinates), which was used to extract street images; and ii) the COVID-19 risk level assigned to each bus stop, which was used as the outcome labels.

We downloaded the images from Google Street View through the application programming interface (API). We used the Python libraries *google_streetview.api* and *request*. For each bus stop (i.e., for each latitude and longitude coordinate), the API request specified the heading parameter = [0, 90, 180, 270]; in addition, we downloaded one image for which the heading parameter was set at default. Consequently, for each bus stop we had five images in total. The images were downloaded with size 640x640 pixels. In this feasibility study there were 1,788 bus stops, and because we had five images per bus stop, we used 8,940 (1,788 x 5) images in total.

Labels (outcome)

We used the bus stop classification released on 2021-05-24,^{2, 3} by the local transport authority in Lima, Peru.² They classify the bus stops in four categories of COVID-19 risk: moderate, high, very high and extreme risk.² Details on how they made this classification are not available. Nevertheless, because this is official information released by a public authority to inform the general population, we trust their classification is based on the best available evidence. In this feasibility work we only used bus stops labelled as *moderate* (n=1,173) and *extreme* (n=615) COVID-19 risk. There were 1,788 (1,173 + 615) bus stops in total. The list of labels was appended five times, so that we would have as many images as labels (NB: we had five images per bus stop as described above).

Class imbalance

There were two outcomes of interest: moderate and extreme risk. However, there were more observations in the moderate category than in the extreme category. That is, there was class imbalance. After splitting the data into the training, test and validation sets, we corrected for class imbalance. We randomly increased the number of observations in the extreme category by 90% in the training dataset (not in validation and test datasets). The original (before correction for class imbalance) training set had 3,519 observations in the moderate category and 1,845 in the extreme category (3,519 + 1,845 = 5,364). After correcting for class imbalance as described before, the training dataset had 3,519 observations in the moderate category (this number did not change) and 3,505 (~1.9 x 1,845) observations in the extreme category (3,519 + 3,505 = 7,024 total sample in the training dataset).

Image data generator

We constructed a dataframe with two columns: the path to each image (i.e., to the exact location where the images were saved) and the corresponding label for each image. This dataframe was passed to the *ImageDataGenerator* function of the *keras.preprocessing.function* library. At this point, we also re-scaled the images between 0 and 1 by dividing by 255. Then, we created three image iterators: one for the training, validation and test datasets (*train_datagen.flow_from_dataframe* function). To the *train_datagen.flow_from_dataframe* function we passed the dataframe with the location of the images and their labels, specified this was a categorical classification problem, a batch size of 32, and a

specific image size for each candidate model (see table below); in addition, the image iterator for the training dataset had the shuffle parameter as *True*.

Convolutional neural networks

We decided to train five candidate convolutional neural networks with the following specifications.

	NASNetLarge	InceptionResNetV2	Xception	ResNet152V2	ResNet101V2
Image size	331 x 331	299 x 299		224 x 224	
Pre-trained weights	ImageNet				
Top layer included?	No				
Trainable parameters	None				
Additional layers	Dense layer with 2 neurons (for the two outcome labels), with <i>softmax</i> activation				
Number of epochs for training	25				
Optimizer	SGD(learning_rate = 0.1, momentum = 0.9, decay = 0.1/number of epochs, nesterov = True)				
Loss function	Binary crossentropy				

SDG: stochastic descent gradient.

In addition, we monitored the validation loss: when the validation loss would not improve in one epoch, then the learning rate was multiplied by 0.1. We also specified an early stop: when the validation loss would not improve for ten epochs, the training would stop. To choose between the five candidate models we did not implement any data augmentation methods.

We chose the model with the best performance (Table 1 in the main text), which was further tuned (Table 2 in main text).

Model Architecture

The chosen model (NASNetLarge) presented in this article is a state-of-the-art convolutional neural network (CNN) pre-trained with the ImageNet dataset of images. The NASNetLarge neural network is widely used in computer vision. It is composed of several layers of convolutional cells that extract the most important features from each image, learning which features are the most characteristic from each image category. Most pre-trained state-of-the-art CNNs, such as the AlexNet or the ResNet50, base their innovations in the use of complex activation functions, efficient designs and special layers, such as the Batch Normalization,⁴ which not only improve accuracy performance, but also reduce the time needed to train a model. The ResNet50 neural network, for example, introduces the concept of residual neural networks, layers that skip their connections to other layers by adding connection shortcuts, thus reducing backpropagation training time and better generalizing models.

1
2
3 However, most of these CNNs had complex design that was built by using trial-and-error methods,
4 oblivious to modern state-of-the-art optimization methods such as the use of evolutionary and nature-
5 inspired algorithms or Reinforcement Learning (RL). This is where Neural Search Architecture (NAS)
6 Networks come in play.⁵ Dubbed as a breakthrough in machine learning automation, NAS networks
7 use *AI for AI*. The network's whole architecture is designed by optimization algorithms, such as
8 Gradient-based search and RL. The idea behind this is that, setting the right restrictions to avoid
9 repetitive inclusion of layers, the network cannot only be trained by its parameters, but by its own
10 architecture, adding and changing layers, activations functions and connections.
11
12
13
14
15
16
17
18

19 Our research uses the NASNetLarge model available by Keras,¹ pre-trained on the ImageNet dataset
20 on 1000 categories. Because we only have two categories to train on (moderate and extreme), we
21 started by replacing the network's fully connected classification layer by a 2-neurons layer, with a
22 softmax activation function.¹ The rest of the original NASNetLarge network was kept unmodified to
23 preserve its proven architecture.
24
25
26
27
28

29 **Model Training**

30 To train our model we used transfer learning. Based on the pre-trained NASNet model, we retained its
31 parameters and froze them, keeping just trainable the last fully connected layer, initializing its
32 parameters randomly by He uniform initialization.⁶ This single-stage training took advantage of the
33 pre-trained parameters speeding up training by just focusing on the last decision layer. The model
34 was trained for 30 epochs. Each epoch is understood as a complete training cycle through the whole
35 train dataset. Data is then fed by batches; once all batches are loaded, an epoch is finished. By using
36 a batch size of 32, training and testing sets are fed to the neural network. The loss function, as we are
37 dealing with a binary classification model, is binary cross entropy.
38
39
40
41
42
43
44
45
46
47

48 As optimizer we used the Stochastic Gradient Descent (SGD).⁷ One of the key advantages of this
49 optimizer is due to its stochasticity in selecting each batch for the backpropagation step. Although the
50 model takes longer to converge, unlike other optimizers, the SGD has been proven to converge better
51 local optima, searching for global optima with much more easiness. This factor helps also to reduce
52 overfitting.
53
54
55
56
57
58
59
60

Activation Heatmap by GradCam Visualization

Model interpretability is a feature that has gained importance since the appearance of methods such as GradCam.⁸ This technique uses the values from the gradients in the model's final feature layer to produce visual explanations, highlighting regions of importance taken by the model to infer a given input. In other words, this technique informs which areas of the input figure were more relevant to make the final classification.

Regions that represent higher gradient values, accounting for most of the last layer activations and network's decision, are represented by being coloured closer to the red portion of the spectrum. By comparison, regions with the lowest activations, not adding much information to the network's final decision, appear as areas closer to the blue portion of the spectrum. These gradient values are taken from the network's last convolutional layer, right before the last pooling layer.

References

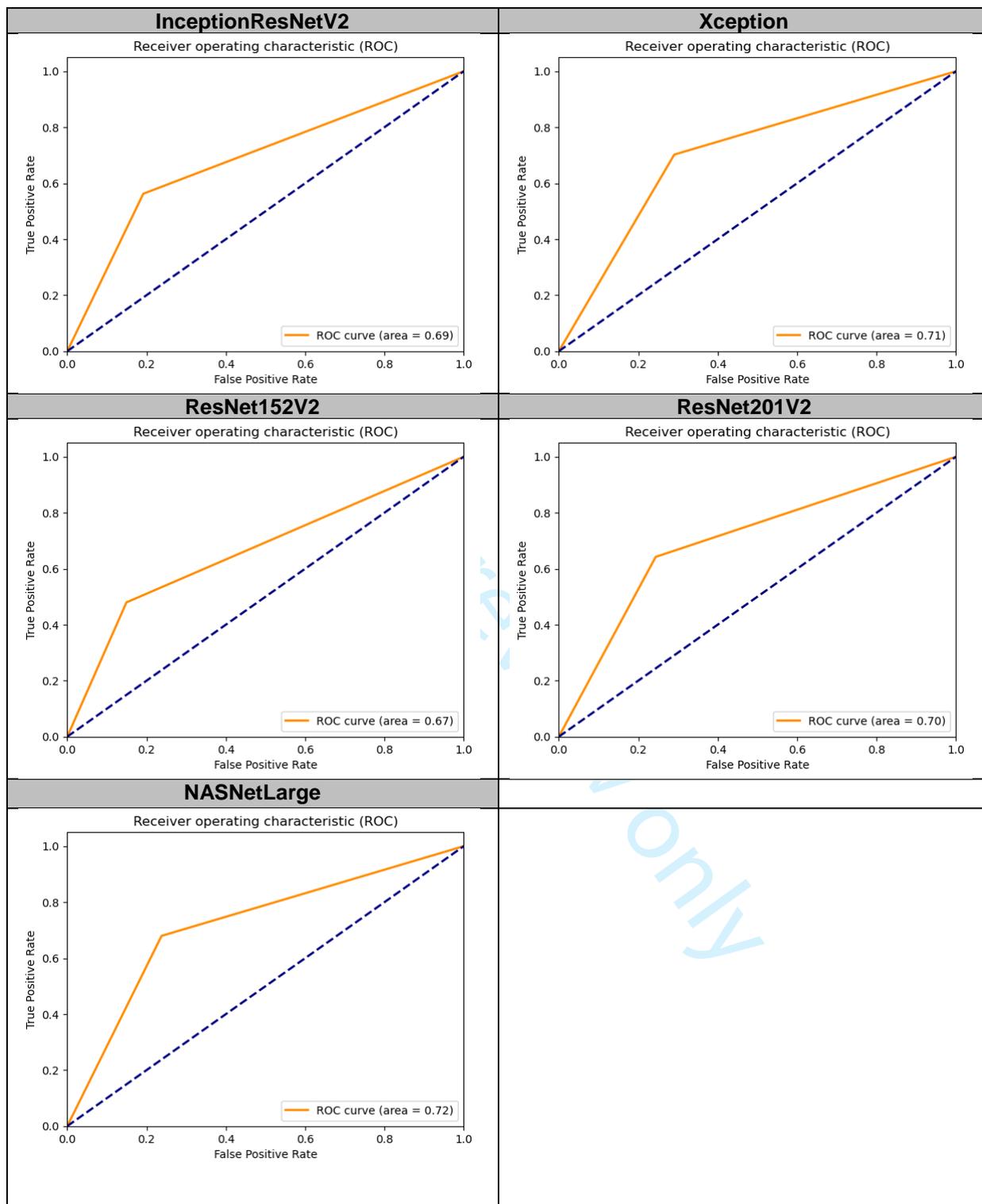
1. Keras applications. URL: <https://keras.io/api/applications/>.
2. Autoridad de Transporte Urbano para Lima y Callao (ATU). Paraderos con Riesgo de COVID - 19. URL: <https://sistemas.atu.gob.pe/paraderosCOVID>.
3. URL: <https://github.com/jmcastagnetto/lima-atu-covid19-paraderos>.
4. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv e-prints* 2015: arXiv:1502.03167.
5. Kyriakides G, Margaritis K. An Introduction to Neural Architecture Search for Convolutional Networks. *arXiv e-prints* 2020: arXiv:2005.11074.
6. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *arXiv e-prints* 2015: arXiv:1502.01852.
7. Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv e-prints* 2016: arXiv:1608.03983.
8. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision* 2020; **128**(2): 336-59.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

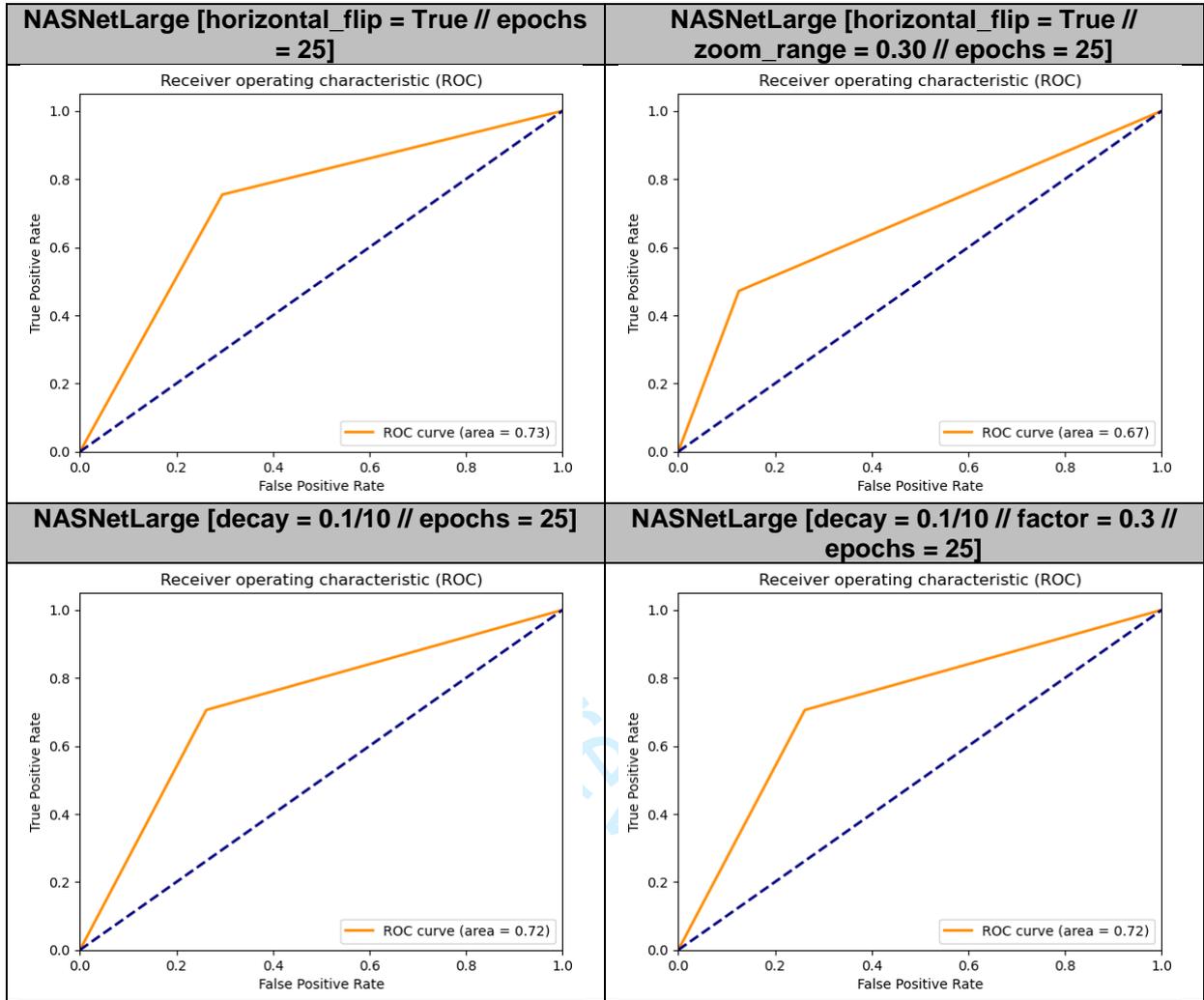
Enseignement Supérieur (ABES) .
Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Receiver Operating Characteristic (ROC) Curves I



Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Ensignement Supérieur (ABES).

Receiver Operating Characteristic (ROC) Curves II



Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies. Enseignement Supérieur (ABES).

only