Analysis of LiDAR Images with the Health Deprivation Index

Author: Florence Sun

1. Introduction

LiDAR (light detection and ranging) is a modern technique for creating high-resolution ground elevation models. It is frequently used in a variety of fields, including geography and forestry [1]. The aim of the deprivation index is to characterize and highlight deprivation at the small-area level, which is meaningful for regional resource planning in the health and social services system. The most frequent used index is multidimensional deprivation index which contains income, work, health deprivation and disability, education, crime, impediments to housing and services, and living environment [2]. The following sections of this paper examine the relationship between LiDAR images and the health deprivation index:

- Relation between LiDAR images and health deprivation index (Section 2)
- Base Model using VGG16 (Section 3.1)
- Base Model using ResNet50V2 (Section 3.2)
- Model comparison and evaluation (Section 3.3)
- Visualization using Grad-CAM (Section 4)
- Discussion (Section 5)
- Multimodal learning (Section 6)

2. LiDAR application in health deprivation index

LiDAR technology can be used to create detailed three-dimensional landscape elevation maps. Several research papers leveraging deep learning approaches to interpret LiDAR pictures have been published in the application of multidimensional deprivation index prediction.

Esra Suel et al. have developed two deep learning-based approaches for quantifying urban inequities such as income, overcrowding, and environmental deprivation in London by combining LiDAR pictures and street level imagery. They discover that using LiDAR photos reduces the mean absolute error, and that the forecast resolution is only limited by the resolution of the input satellite imagery [3].

Matthew et. al. proposed a series of elevation embeddings by using unsupervised deep learning methods to predict seven English indices of deprivation (2019) for small geographies in the Greater London area. They discover that their strategy improves Root-Mean-Squared-Error by up to 21% when compared to using typical demographic features alone [4].

Deep learning algorithms have also been shown to extract key features from LiDAR or satellite data for archaeological feature detection [5] and poverty prediction [6] in other studies.

As a result, it is reasonable to conclude that deep learning approaches can extract essential aspects from LiDAR images, such as building features and grassland, in order to forecast the health deprivation index. In the meanwhile, it suggests that the image resolution may restrict the forecast accuracy.

3. Transfer Learning

Transfer learning is a machine learning method that focuses on storing and transferring knowledge learned while addressing one problem to a different but related problem, which is commonly used in convolutional neural network models.

For transfer learning purposes, two models from distinct CNN families are chosen as the base model in this section. The first is VGG16, and the second is ResNet50V2.

However, the pre-trained model should be contextually similar to the target domain for transfer learning. This is especially difficult with LiDAR photographs since VGG and ResNet are trained on the ImageNet dataset, which contains item collections such as animals, automobiles, and buildings, rather than LiDAR photos. As a result, fine-tuning the base model is required, which means that some VGG and ResNet layers must be set as trainable for learning purposes.

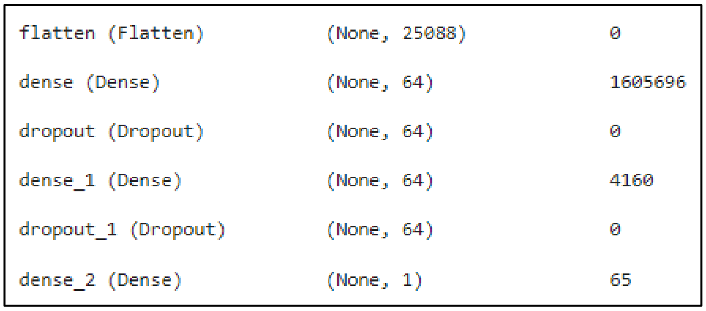
3.1 Base Model Using VGG16

3.1.1 Model Architecture

In ImageNet, VGG16 is a convolutional neural network that achieves 92.7 percent top-5 test accuracy. The input to the cov1 layer of the VGG16 model is a 224*224 RGB picture. Five blocks are used to pass the image, each of which contains two or three convolutional layers, as well as a maximum pooling layer. The majority of filters are 3 by 3 pixels in size. For 3*3 convolutional layers, the convolutional stride and spatial padding are both set to 1 pixel, ensuring that the spatial resolution is kept after convolution. Max pooling is done with stride 2 over a 2*2 pixel window. There are three identical fully-connected layers after five blocks. The soft-max layer is the final layer. The health deprivation index, on the other hand, is a continuous response. As a result, the last layer is no longer included in the new model [7].

Another six layers are built on top of VGG16, as seen in table 1. The flatten layer is the first. The dense layer, which has 64 units and a Relu activation function, is the second. Relu activation function is to convert all negative values to 0. The third layer is a dropout layer, which prevents overfitting by randomly setting input units to 0 with a frequency of 0.5 in this case at each step during training period. The fourth and fifth layers, again with the identical settings, are dense and dropout layers. The last layer is a dense one-output unit layer with a linear activation function.

Table 1 Added layers on top of VGG16



```
flatten (Flatten)          (None, 25088)          0

dense (Dense)              (None, 64)             1605696

dropout (Dropout)          (None, 64)             0

dense_1 (Dense)            (None, 64)             4160

dropout_1 (Dropout)        (None, 64)             0

dense_2 (Dense)            (None, 1)              65
```

3.1.2 Model Training

VGG architecture is based on the ImageNet dataset, which includes animal, vehicle, and building collections rather than LiDAR photographs. As a result, the underlying model must be fine-tuned, which requires several VGG layers as trainable.

Figure 1 depicts the training and validation loss over a period of ten epochs. With each epoch increase, the training loss decreases from 1.0 to roughly 0.6. When the epoch is 7, the validation loss falls steadily until it approaches the minimum state. The validation loss converges after epoch 7.

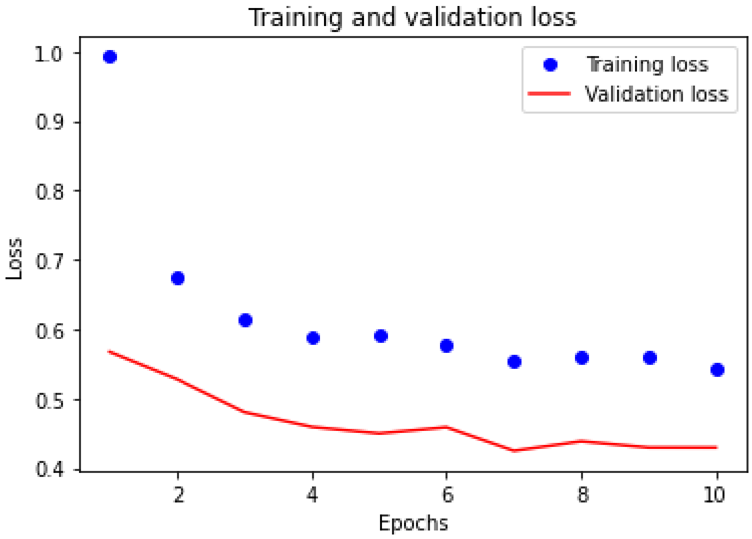


Figure 1. Training and validation loss from different epochs.

Therefore, after 7 epochs of training, the model achieves the best prediction performance. The weights of the top performing model are saved to the checkpoint folder using the callback method.

3.1.3 Test data Prediction

The dataset is divided into two parts: training and testing. To estimate weights, training data is fed into the model, while the model prediction is evaluated using test data. The model's prediction performance is measured using two metrics. For the test dataset, the mean squared error (MSE) is 0.446. The R2 score, on the other hand, describes how much of the variance in the dependent variable can be predicted by the independent variables. is 14.36%.

3.2 Base Model using ResNet50V2

3.2.1 Model Architecture

The ResNet-50 model is divided into five stages, each with its own convolution and identity block. There are three convolution layers in each convolution block, and three convolution layers in each identity block. Around 23 million trainable parameters exist in the ResNet-50 [8].

Another six layers are placed on top of the ResNet base model, which are the same as the top six layers of the VGG16-based model seen in section 3.1.1.

3.2.2 Model Training

As LiDAR imagery is not included in the ImageNet database, the layers of the ResNet base model are also set as trainable, similar to the VGG16 based model.

The model was trained for ten epochs, but it was terminated after epoch 8 because the validation loss converges afterwards. Figure 2 clearly indicates that training loss continues to decrease after epoch 6 and then stops changing. At epoch 5, the validation loss approaches a minimum of roughly 0.40.

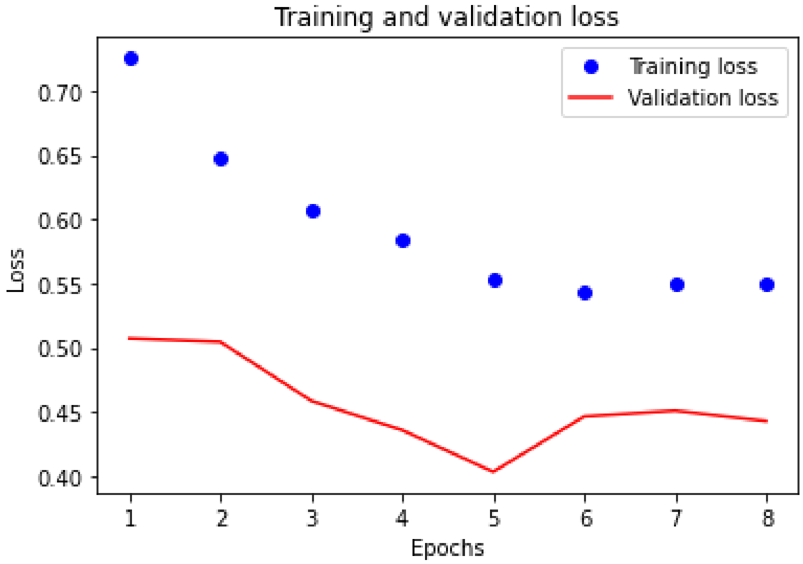


Figure 2 Training and validation loss for 8 epochs

3.2.3 Test Prediction

The model's prediction performance is assessed using the MSE and R2 scores. The MSE is 0.418, and the R2 value is 19.73%, indicating that LiDAR image characteristics can explain 19.73% of the variance in the health deprivation index.

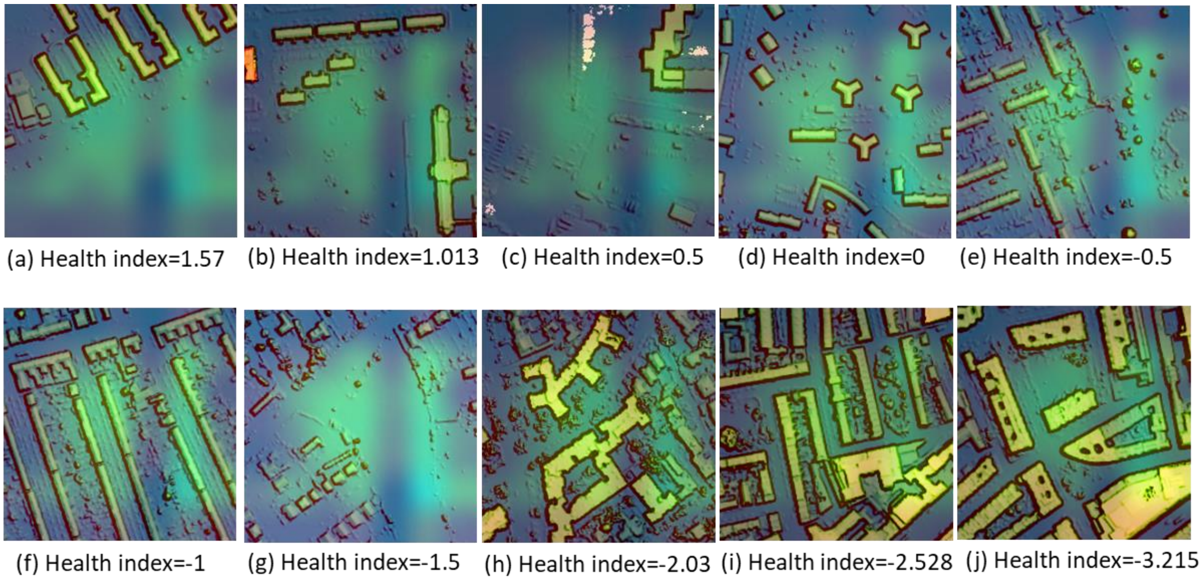
3.3 Model Comparison

The base model employing ResNet50V2 obviously beats the model based on VGG16 in terms of predictive power, as it reduces MSE by 6.27% and improves R2 score by 37%. Because of the differences in architecture, ResNet50V2 takes less time to train the model than VGG16 in terms of computational time [8].

4. Visualization using Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) uses the gradients of target flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image [9].

In this section, ten health deprivation indices ranging from -3.215 to 1.57 with a step of roughly 0.5 are chosen to represent the entire health variance. The CNN model based on ResNet50V2 is then visualized using one image for each health deprivation measure. Figure 3 shows the heatmap and its projection on LiDAR images for ten health deprivation indexes.


(a) Health index=1.57 (b) Health index=1.013 (c) Health index=0.5 (d) Health index=0 (e) Health index=-0.5

(f) Health index=-1 (g) Health index=-1.5 (h) Health index=-2.03 (i) Health index=-2.528 (j) Health index=-3.215

Figure 3. Heatmap and its projection on the image for ten health deprivation indices.

The results above show the features the ResNet model detects and how it uses them to predict the health deprivation index. The health index of 1.57 is shown in Figure 3(a). The highlighted areas demonstrate that the model not only picks up the wilderness but also portions of the long rectangle blocks. As a result, the predicted value is roughly -0.4, which is significantly lower than the true value.

The health index of 1.013 is depicted in Figure 3(b). In a similar way to figure 3(a), the model captures the large lengthy block on the right side of the image and generates a forecast of roughly -0.65.

Figure 3(c) shows a health deprivation index of 0.5, indicating that the model makes advantage of the flat terrain on the left as well as some blocks in the upper right image. The ResNet model only captures a small part of blocks and consequently the prediction is around -0.2.

The health deprivation index of 0 is shown in Figure 3(d). The test prediction is about 0.01-0.02 since the model catches the most significant aspects in the image, demonstrating the high test accuracy for the health index of 0.

Figure 3(e) shows the health index of -0.5. The forecast value for true value is similarly around -0.5. The model's excellent accuracy is owing to the fact that it successfully extracts not only the blocks that represent housing but also some clusters that resemble grassland.

The LiDAR image corresponding to a health index of -1 is shown in Figure 3(f). The prediction, however, turns out to be between -0.8 and -0.6, which is higher than the true value. The reason is that the model is beginning to show another flaw of only catching a portion of the blocks rather than all of them

Figures 3(g)–(j) are similar to those in figure 3(f). The ResNet model recognizes the major large blocks in LiDAR imagery.

To summarize, the ResNet model learns from LiDAR pictures that block number, block size, and flat ground are key properties, and it uses these features to predict the health deprivation index, as demonstrated by the Grad-CAM graphic.

5. Discussion

5.1 Limitation of LiDAR images

The ResNet model's R2 score on the test dataset is around 19%, indicating that it only captures 19% of the variance in response variable with regard to the independent factors. Meanwhile, the Grad-CAM displays crucial features of the model, such as block size and number, as well as plain terrain. It demonstrates that the LiDAR image set can be used to forecast the multidimensional deprivation index.

It does, however, highlight some of the model limitations. It is reasonable to conclude that the multidimensional index is influenced not only by the number and size of building units, but also by the structure and infrastructure, which are not visible in LiDAR images. In other words, the LiDAR images' resolution is insufficient for the model to identify the block difference. As a conclusion, the model's predictive power needs to be improved by increasing the resolution or adding more data.

5.2 Ethical Challenge

The ethical and privacy issue is one of the most urgent challenges in data collection and use. LiDAR technology's visual data has the potential to see private property and gather sensitive personal information. Residents may be unaware that their living conditions or properties have been captured and published over the internet, as the LiDAR image collection used in this work contains geographic information. They are unlikely to take action quickly even if they are aware that their privacy rights have been violated. In the meantime, individuals have few options for providing informed consent for this geospatial data [10].

Another issue is the application of geospatial data. Many countries have disparate data protection and storage requirements, which could be an issue if data is transferred in areas where human rights and humanitarian norms are lacking. Sensitive geographical data is also released as a result of unsecured mobile networks or platforms [11].

Surveillance and privacy issues can be addressed by de-identifying geographical images. As a result, if the resolution is low enough or the distance is great enough, no individual information can be detected, these hazards can be avoided. The dataset's purpose, on the other hand, determines how low the image resolution can be.

The potential solution for the data usage loophole is to legislate applicable laws or regulations governing data usage and limit users' access to sensitive data.

6. Multimodal Learning

Multimodal learning requires collecting data from several sources. Unstructured data such as photographs and videos, for example, are combined with structured data to boost the model's predictive power. It is obvious from the previous section that LiDAR imagery has limits in terms of prediction. The LiDAR photos were taken in several parts of the United Kingdom, each with its own area code. Therefore, the area code information is used as the structural data.

Multimodal learning is divided into two portions in this paper, as indicated in Figure 4. The structural data is fed into one neural networking model, while the LiDAR images are fed into ResNet50v2 CNN model. After that, two models are concatenated.

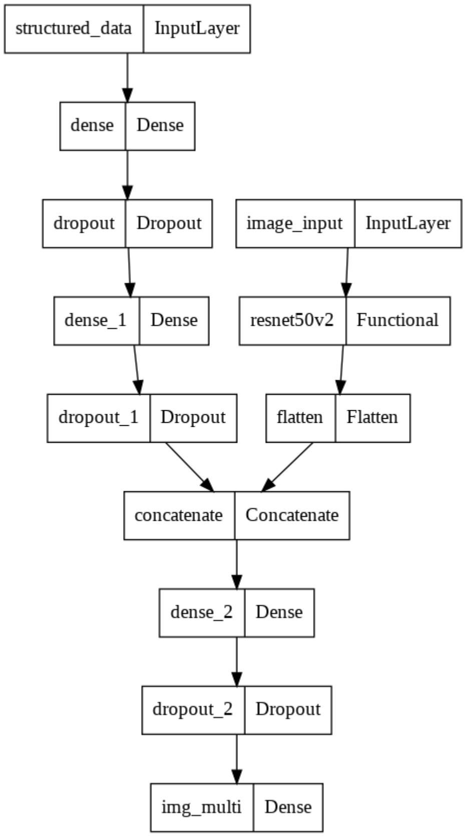


Figure 4. Multimodal learning architecture.

Figure 5 depicts the training and validation losses. It shows that the model reaches its peak performance at epoch 6, with a validation loss of roughly 0.37. The validation loss begins to rise after that, indicating overfitting.

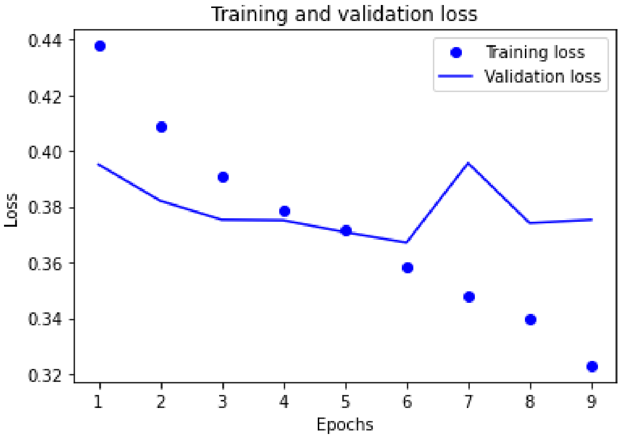


Figure 5. Training and validation loss for multimodal learning.

The R2 score and MSE for three models are shown in Table 2. With the best R2 score and lowest MSE, multimodal learning outperforms the other two models.

Table 2. R2 score and MSE for three models

Performance Metric	Image_only_VGG16	Image_only_ ResNet50v2	Multimodal Learning
R2 Score	14.36%	19.73%	29.83%
MSE	0.446	0.418	0.367

References

[1] Shaker, A., Yan, W.Y., & LaRocque, P.E., (2019). Automatic land-water classification using multispectral airborne LiDAR data for near-shore and river environments. *ISPRS Journal of Photogrammetry and Remote Sensing 152, 94–108*.

[2] Atkinson, A. B. (2021). Multidimensional deprivation: Contrasting social welfare and counting approaches - *The Journal of Economic Inequality 1, 51-65*.

[3] Suel, E., Bhatt S., Brauer M., Flaxman S., & Ezzati M., (2021). Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. *Remote Sensing of Environment 257, 112339.*

[4] Stevenson M., Mues C., & Bravo C., (2021). Deep residential representations: Using unsupervised learning to unlock elevation data for geo-demographic prediction. *arXiv:2112.01421*

[5] Albercht C., Fisher C., Freitag M., Hamann H., Pankanti S., Pezzutti F., & Rossi F., (2019). Learning and Recognizing Archeological Features from LiDAR Data. *arXiv:2004.02099v1*

[6] Jean N., Burke M., Xie M., Davis M., Lobell D., & Ermon S., (2016). Combining satellite imagery and machine learning to predict poverty. *Science 353: 790-794.*

[7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).*

[9] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. *In Proceedings of the IEEE international conference on computer vision (pp. 618-626).*

[10] Berman, G., de la Rosa, S., & Accone, T. (2018). Ethical considerations when using geospatial technologies for evidence generation.

[11] Slonecker, E. T., Shaw, D. M., & Lillesand, T. M. (1998). Emerging legal and ethical issues in advanced remote sensing technology. *Photogrammetric engineering and remote sensing, 64(6), 589-595.*