Astin Bulletin (2024), **54**, pp. 280–309
doi:10.1017/asb.2024.9

International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

A representation-learning approach for insurance pricing with images

Christopher Blier-Wong[1], Luc Lamontagne[2] and Etienne Marceau[3]

[1]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada, [2]Département d'informatique et de génie logiciel, Université Laval, Québec, QC, Canada and [3]École d'actuariat, Université Laval, Québec, QC, Canada
Corresponding author: Christopher Blier-Wong; Email: cblierwo@uwaterloo.ca

Received: 12 May 2023; **Revised:** 23 December 2023; **Accepted:** 21 February 2024; **First published online:** 15 March 2024

Keywords Representation learning; insurance pricing; embeddings; image models; unstructured data

Abstract

Unstructured data are a promising new source of information that insurance companies may use to understand their risk portfolio better and improve the customer experience. However, these novel data sources are difficult to incorporate into existing ratemaking frameworks due to the size and format of the unstructured data. This paper proposes a framework to use street view imagery within a generalized linear model. To do so, we use representation learning to extract an embedding vector containing useful information from the image. This embedding is dense and low dimensional, making it appropriate to use within existing ratemaking models. We find that there is useful information included in street view imagery to predict the frequency of claims for certain types of perils. This model can be used as in a ratemaking framework but also opens the door to future empirical research on attempting to extract which characteristics within the image leads to increased or decreased predicted claim frequencies. Throughout, we discuss the practical difficulties (technical and social) of using this type of data for insurance pricing.

1. Introduction

As a novel data source, images can intervene at many places within property and casualty insurance applications. For instance, they may improve the quoting process by filling in some fields in the quoting questionnaire (number of stories, material type of the facade and presence of garage), either by an insurance agent or automatically with an artificial intelligence system. Going a step further, one may also use images directly within a ratemaking model to investigate whether they can predict future claim counts or severity. In a claims management process, images may accelerate the handling procedure: if a customer provides an image of the damaged house or vehicle, one may estimate (manually by a claims adjuster or by an artificial intelligence system) the cost of repair or replacement of the damaged goods; this will lead to closing claims at a faster rate. If done correctly, the above examples enhance/simplify the customer experience, improve actuarial fairness or reduce operating expenses.

For an insurance company, ratemaking plays a crucial role. Traditional ratemaking models rely on structured data such as policyholder demographics, coverage details and loss history. However, the recent surge in the availability of unstructured data, particularly images, has opened up new possibilities for enhancing ratemaking models and improving risk assessment. One research question we seek to answer in this paper is *Are there useful information within street view imagery to predict the frequency or the severity of home insurance claims?* Our main methodological contribution is to propose a framework based on representation learning that will enable us to answer this research question. We use state-of-the-art image recognition techniques to extract meaningful information from images, which we then use to predict the frequency and severity of insurance claims. This significantly advances ratemaking, allowing for more accurate and nuanced risk assessments, resulting in fairer pricing for policyholders.

Concrete examples of uses of images for risk management include Biffis and Chavez (2017), where the authors use satellite data to model weather risk using precipitation variability indices. More recently, in an actuarial context, Zhu (2023) combines economic and weather data to predict a production index for crop yield. In Brock Porth *et al.* (2020), the authors use pasture production indices derived from a satellite-based sensor to predict crop yield. Meanwhile, Wüthrich (2017) illustrate how driving styles can be summarized in an image through velocity-acceleration heatmaps. These heatmaps were further used for risk classification; see, for instance, Gao and Wüthrich (2019), Zhu and Wüthrich (2021). From a claims management perspective, the authors of Doshi *et al.* (2023) predict the cost of repairing a car after an accident when provided an image of a damaged vehicle. Street view imagery (SVI) has proven useful for many applications; see Biljecki and Ito (2021) for a review in urban analytics and geographic information science. As an example, Fang *et al.* (2021) use SVI to produce land-use classification and land-use mapping, critical tools for urban planning and environmental research. In Chen *et al.* (2022) and Blanc (2022), the authors use SVI to perform information extraction from images to automate the data-collection process, improving the quoting process for customers. To the best of our knowledge, the current work is the first to use images to predict claim frequency or claim severity. Such a study is now feasible due to recent advances in image models, their performance for practical situations and the interest of insurance companies to stay ahead of the competition by adopting data-driven strategies in ratemaking and risk selection.

In this paper, we propose a framework that uses images as inputs to ratemaking models. One problem we will face is that insurance data has a low signal-to-noise ratio (Wüthrich and Ziegel, 2023), and insurance datasets typically do not have millions of examples as in image models. Further, the state-of-the-art image models we will use in this paper contain tens of millions of parameters. For this reason, we will not be able to use state-of-the-art image models directly since these will almost certainly overfit the training data. Instead, our framework is based on representation learning, where we will use image models pre-trained on a large dataset and fine-tune them to a dataset related to household information. A secondary goal of this paper is to perform an empirical study to determine if there is useful information within SVI to predict home insurance losses. We will show that every method we propose to incorporate images in the ratemaking model improves the prediction of claim frequency for the sewer backup and water damage perils.

This work fits within the machine learning literature in actuarial science; see Richman (2020a,b) or Blier-Wong *et al.* (2021b) for early reviews. In particular, we use representation learning to determine if there is useful information in images for insurance pricing. For an overview of representation learning from an actuarial perspective, we refer the reader to Blier-Wong *et al.* (2021a). In particular, Richman (2020a) have advocated for using entity embeddings to transform categorical data into dense vectors; this idea was also suggested in Blier-Wong *et al.* (2021a), Shi and Shi (2022), Embrechts and Wüthrich (2022), Wüthrich and Merz (2023). In Delong and Kozak (2023), the authors suggest initializing the parameters associated with the entity embeddings by first training an autoencoder. In Avanzi *et al.* (2023), the authors propose a model called GLMMNet to include mixed effects within the embedding layer.

What sets our approach apart from those proposed in the existing literature is that we train our embeddings in an unsupervised way with respect to our prediction task of interest, meaning that we do not use the insurance loss data to construct the representations. We will adapt the predictive learning framework proposed in Blier-Wong *et al.* (2021a) for a ratemaking model that uses image data as input. This framework lets one combine traditional actuarial variables with emerging variables, such as spatial, image and textual data, within a simple predictive model (such as a GLM) for insurance pricing. In the proposed ratemaking framework, one uses representation learning for each source of emerging variables; this step has the dual purpose of extracting non-linear transformations and interactions within each emerging variable and of converting the (potentially unstructured) source of data into a dense vector of numerical values. Then, one combines each source of emerging data along with the traditional variables into a vector of features, which will be the input to a regression model. Since, through representation

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

learning, most useful non-linear transformations and interactions are already extracted, one may use a simple model like a GLM to predict the claim frequency or severity. Applications of the unsupervised representation learning framework in actuarial science include Blier-Wong *et al*. (2022), which uses census data organized in a spatial way to construct geographic embeddings of postal codes to predict homeowners insurance claim frequency and Xu *et al.* (2022), who use pre-trained BERT models to construct representations of textual descriptions of claim data to predict the expected mean payment for truck warranty. The advantage of unsupervised representation learning is that we may construct complicated models for training the representations using large datasets of high-dimensional information like textual, image and spatial data and that these models do not overfit the task of interest, which in our case refers to predicting the future claim frequency/severity for an insurance contract.

The remainder of this paper is structured as follows. We present the general framework for insurance ratemaking using unsupervised embeddings in Section 2. In Section 3, we present the SVI data that we will use to construct the representations in this paper, including the steps we take to prepare and clean the data. In Section 4, we construct representations of images starting from pre-trained image models. We start with a construction method that does not involve much effort and gradually increases the flexibility to determine when the model is useful enough for practical uses. In Section 5, we use the embeddings constructed in the previous section to construct frequency and severity models using a real insurance dataset and we conclude in Section 6.

2. Unsupervised representation learning framework

A key step of our framework is to project SVI data into a low-dimensional vectorial format that captures useful features for insurance applications. In this section, we explain how one may use external data relevant to insurance to *guide* a model to capture the useful parts of an image and let go of useless information. Word embeddings inspire the general approach in natural language processing: instead of using one-hot encodings of words; it may be more convenient to use word representations that capture syntactic and semantic word relationships (Turian *et al.*, 2010; Mikolov *et al.,* 2013; Bengio *et al.,* 2014). The general framework for unsupervised representation-learning in an actuarial context is described in detail in Blier-Wong *et al*. (2021a). In particular, it contains a general discussion of the data types, the intuition behind the representation learning framework and examples of applications in actuarial science. That framework has been applied to spatial data in Blier-Wong *et al.* (2022). Also, the authors of Lee *et al.* (2020) and Xu *et al.* (2022) use a framework that can be considered as a special case of the one described in Blier-Wong *et al*. (2021a) with textual data.

As explained in Blier-Wong *et al.* (2021a), we believe that the unsupervised representation learning approach is well suited for insurance pricing due to the nature of insurance data (limited number of observations and low signal-to-noise ratio). However, from our experience, it only performs well if the resulting representation is related to the task of interest: adapting the embeddings to the insurance domain ensures that the insights generated by the models are meaningful and actionable for risk analysis. For this reason, we train our unsupervised embeddings on a task similar to our regression task of interest. Selecting appropriate related tasks is a vital step in constructing useful embeddings. Incorporating relevant information in the modelling process results in more accurate risk assessments and pricing, ultimately leading to fairer premiums for policyholders and better financial stability for insurers.

An intuitive interpretation of the framework is that we aim to construct proxies of insurance-related information from SVI such that these proxies will be useful for ratemaking. These proxies are intermediate representations in a large image model trained on insurance-related tasks. The closer the insurance-related tasks are to predicting claim frequency or severity, the more useful the proxies should be. The representations should satisfy every principle related to rating variables; in particular, they should not rely on any protected attribute, such as the race or gender of people appearing in the SVI.

Our approach is in the view of few-shot learning or unsupervised multitask learning in natural language processing (Radford *et al.,* 2019; Brown *et al*., 2020) or image classification (Tian *et al*., 2020).

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

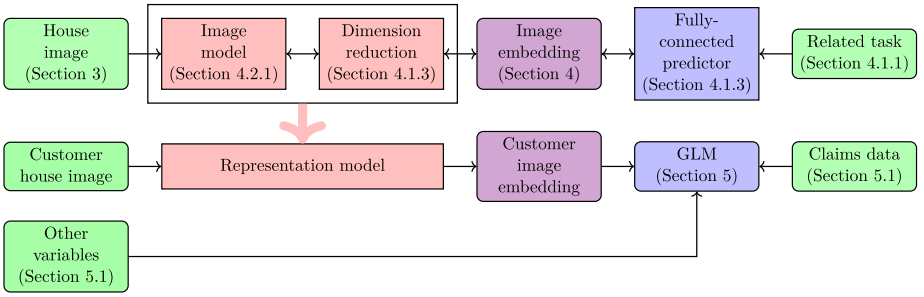

Figure 1. *Framework for the representation-learning framework.*

In few-shot learning, one attempts to design machine learning models that can effectively learn valuable information from a small amount of data and make accurate predictions or classifications. Few-shot learners sometimes start with task-agnostic or general multi-task methods, which are not tailored to any specific task. They are designed to capture broad patterns and information across various data types. Following this initial phase, they may undergo a process of fine-tuning, where they are specifically adjusted and optimized for more relevant and targeted tasks. This two-step process ensures that the model is easily adaptable to new tasks and can easily be guided to perform well on new specific tasks of interest. In an image classification context, the authors of Tian *et al*. (2020) suggest that finding a good starting representation, followed by a linear classifier, outperforms other state-of-the-art few-shot learning models. In our case, we pre-train an initial representation model on auxiliary tasks related to insurance in a data-rich and high signal-to-noise context using data from municipal evaluations. We then transfer knowledge from the initial auxiliary tasks to a new, data-scarce and low signal-to-noise ratio model to predict claim frequency or severity.

We summarize the framework used for our experiments in Figure 1. The black arrows in Figure 1 indicate the flow of information: since there is a path between the related tasks and the image model, the related task influences the image model. On the other hand, since there is no path from the claims data to the image model, the claims data does not influence the image model. The representation learning framework appears in the first row, while the actuarial pricing model appears in the second and third rows. The arrow in light red corresponds to the knowledge transfer from the representation learning model to the predictive model. The main task on the first row is to train a model that inputs the SVI of a house and outputs predictions on insurance-related tasks. One characteristic of this first model is a bottleneck near the end of the model, called image embedding, such that all the information about the input image is condensed into a small vector that will eventually yield a prediction for the related tasks. Because of the bottleneck, the neural network is implicitly regularized such that it is forced to contain all useful information about the related task and little information about other unrelated tasks. Note that most arrows flow both ways in the first row. The forward direction is straightforward since this model takes a house image as input and a prediction of the related tasks as output. The backward direction means that the parameters of the components (image model, dimension reduction and fully-connected predictor) depend on the related task data since one optimizes the parameters of these models to perform well on the related tasks.

Once the representation-learning step in the first row is completed, we no longer change the parameters of the image model or the dimension reduction models: within our approach, they are considered fixed for the remainder of the framework. We use the trained representation learning model to construct an embedding of the SVI for a potential customer. We then use the customer house image embedding, as well as other variables (traditional actuarial variables or other sources of novel information such as geographic embeddings or textual embeddings) to construct a predictive model for claim count or claim severity. If we construct useful embeddings in the representation-learning step, it is unnecessary to use flexible models for the claim count or claim severity model since all useful non-linear transformations

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

and interactions between parts of the SVI are already captured in the representation-learning step. Note that there are only two-way arrows between the GLM and the claims data since the representation model does not depend on the claim data. It follows that the parameters of the representation model do not contribute to the degrees of freedom in the GLM, so we get the advantages of neural networks (much flexibility to learn useful non-linear transformations and interactions between variables) while still being able to rely on the statistical properties of GLMs through maximum likelihood estimation.

One notices from Figure 1 that the house images for training the representation model do not need to be the same as those for insurance pricing. One could train the representation model using data from a city/state/country and use that same model to construct customer image embeddings for houses in another territory. Assuming that there are no significant changes in how the SVI is collected and assuming the training data has seen a wide variety of house styles and years (mix of high- and low-income neighbourhoods, a mix of houses/condos/apartments), one will be able to generate predictions on new houses that were not used in constructing the representation model.

3. Image data

Most of the applications of images we have encountered in the actuarial science or risk management literature have been dedicated to comparing images before and after an event to examine the status of a property (the occurrence or extent of damages); see the related works in Section 1. What sets our approach apart is that we only use "before" images, that is, an image of the insured property in its normal condition. To the best of our knowledge, we are among the first to use images of undamaged houses as inputs to a ratemaking model. For this reason, we must define the desirable attributes for image data and determine which data cleaning is necessary for ratemaking. In this section, we discuss the data we will use to construct embeddings of house images.

If one uses images as input to a ratemaking model, then the images and their content should be convenient to use and satisfy the properties of rating variables. One place to look for such properties is the Actuarial Standards of Practice (ASOP) from the Actuarial Standards Board (http://www.actuarialstandardsboard.org/). Such standards are guidelines on the applications, methods, procedures and techniques actuaries follow when conducting their professional work in the United States. Relevant ASOPs are No. 12 (Risk Classification), No. 23 (Data Quality) and No. 56 (Modelling). Below, we list some considerations that, in our opinion, the image data should have such that they are convenient to use and respect the ASOPs:

1. Relevance: The image should contain relevant information about the house that could affect insurance rates (irrelevant images could lead to irrelevant rates).
2. Causality: The image should depict elements that could cause or be impacted by losses, ensuring that the image model can capture the components that generate risk.
3. Data quality: The images adequately identify risk factors (for instance, high enough resolution).
4. Representation: The image dataset should contain diverse images and be representative of the population we intend to insure (different types of houses/buildings, different architectural types from many historical periods). They should also be photographed under different weather conditions. Further, one should avoid using a dataset that over or under-represents certain regions. Note that bias could appear in the data if, for instance, the camera quality is higher in wealthier neighbourhoods.
5. Availability: Images are available or quickly obtainable for every (most) potential customer for the desired market.
6. Privacy: The model should not use personally identifiable information or sensitive details that could identify people in the image (for instance, faces of pedestrians, street address and licence plate number).

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

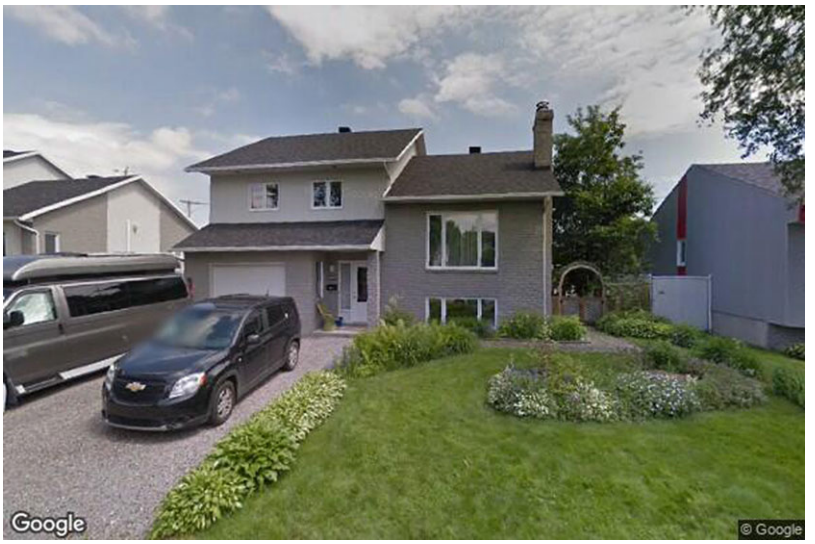

Figure 2. *Ideal candidate for facade image.*

7. Legal compliance: For example, the images should not contain information that reveals protected attributes such as the gender or race of policyholders. If permission to use the image data is required, the insurance company should obtain this permission before beginning the quoting process.
8. Ground truth data: One has access to a dataset of features describing the images such that we may fine-tune the model (only if using the representation-learning approach).

Within the context of this paper, the input data will be SVI, and we henceforth use input data, image data and SVI interchangeably. While many sources of image data were considered to answer our research question (see the discussion for more details), we decided to use data from Google Street View for a few reasons. First, images of many houses are available online; hence there is no need for customers to provide images of their homes during the quoting process. Second, they have an easy-to-use API that can easily be incorporated within a quoting process. The arguments of this API are the coordinates of interest and the requested image size, while the response is a unique JPEG file. Therefore, the availability consideration is satisfied. We note that one can apply the machine learning model, initially trained using data from one specific geographical area, to assess or make predictions about a different geographical area as long as the image dataset satisfies the representation consideration across both places.

We present, in Figure 2, an example of the ideal candidate for images that we wish to include within our study. The house in the image is unobstructed, such that information about, for instance, the roofing quality, the facade material, the number of stories or the presence of a garage. Further, the house is centred within the image and contains contextual information such as parts of neighbours' homes, the presence or absence of trees or electrical lines, etc. Finally, the image does not contain sensitive information that an insurance company does not want to consider within the quoting process. This image, therefore, satisfies the relevance, causality, data quality and privacy attributes.

It is not always true that SVI satisfies most of the desirable properties as in Figure 2. For this reason, one must arrange the dataset before using it for ratemaking purposes. Further, we do not want to do this manually (since we would not want to do this manually within a quoting model in production); we will instead rely on machine learning techniques. We distinguish between two types of cleaning. The first is unsalvageable data, that is, containing no information about the dwelling; in that case, we will need to filter out that data and discard that observation since the image does not satisfy the relevance or causality criteria. We present examples of such scenarios in Figure 3. These situations happen when

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Figure 3. *Examples of images requiring filtering.*

Figure 4. *Examples of images requiring censoring.*

Figure 5. *Steps in image cleanup.*

an object hides the house of interest (for instance, when a truck is parked in front of the house and completely obstructs the house) or when the house is on a large piece of land (in which case, we may only see the entrance of the property and not the house itself). Other examples are that there is no house in the picture, the image is not available in Google Street View, and the homeowner requested Google to remove the house from Street View (in which case the house will be blurred).

The second is data which may contain information that should not be used for insurance pricing, which may fail the privacy and legal compliance attributes. In these cases, we could remove the segments of the image which cause problems. For instance, items in front of the house temporarily (waste bins, debris, etc.) should not be considered within the image since they do not cause or are not impacted by insurance risks. Such examples are provided in Figure 4.

To identify the cases where we must clean or censor the image, we rely on image segmentation methods. Following Blanc (2022) and Blier-Wong *et al.* (2021a), we use the pre-trained semantic segmentation models from Zhou *et al.* (2017) and Zhou *et al.* (2019) with the `ResNet50dilated + PPM_deepsup` models to obtain the categories of objects in the images. We present an example image in Figure 5 with its semantic segmentation (first and second panes). One may compute the percentage of the image that is a house, building or edifice from the mask of these categories (third pane of Figure 5). Within our experiment, if less than 5% of the image is of these categories, we will assume that there is no information about the house to be useful within our application and discard this image. Doing this step removes 4.134% of the images in our dataset, which means that in practice, the insurance company could use this ratemaking framework for over 95% of citizens in the city of Québec.

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

To mask humans and potentially temporary objects, we obtain the segmentation from the images, then construct a mask for the categories corresponding to person/individual and to a list of objects potentially destined for garbage collection such as seat/desk/lamp/toy/pillow (fourth pane of Figure 5). We then replace these pixels with the average pixel value to hide the problematic parts of the image (fifth pane of Figure 5).

4. Representation of related tasks from street view imagery

In this section, we will apply the framework outlined in the first row of Figure 1 to the dataset of SVI presented in Section 3. We will first present the data we selected for the related tasks, that is, the data used to construct the representation learning model. We then present the architecture for the representation models and explain how to extract the embedding vector from the representation learning model. Finally, we will present the training strategy and some information extraction results.

4.1. Constructing representations

4.1.1. Related task data

Most readily available image models are pre-trained on ImageNet (Deng *et al*., 2009), a widely used dataset containing labelled images that serve as the foundation for training deep neural networks in object recognition, containing over 20,000 classes such as banana, banjo and baseball. Starting with pre-trained image models to train our image model is useful since the early layers of these models learn convolutional filters that identify different patterns in the image. However, the feature map they provide may not be useful for insurance applications. For this reason, we must fine-tune the model to a feature space which may be more related to houses since this is the insured good we are considering within our insurance application.

One may use many types of information to construct embeddings. However, choosing a related task that is close or similar to our eventual task of interest (predicting future claim frequency or severity) will yield embeddings that are more useful for that task of interest. Another consideration is the availability of the data for the related task. Ideally, the data will have less irreducible uncertainty than insurance loss data and have a high number of observations such that the representation model can learn flexible features. Another aspect of the related task is that there are many of them, such that the intermediate representation does not overfit on one particular aspect of the image but on the general contents of the image that may be related to insurance losses. Within the context of our representation-learning framework, the actuarial priority lies in selecting the appropriate related tasks. One needs a good understanding of risk factors and the insured product to select relevant tasks that yield useful transferable representations.

Within this paper, we use property assessment data, which is public data collected by the city, to determine the property taxes for a building. These assessments are conducted by an evaluator, whose roles are to keep an inventory of buildings, establish the value of these buildings and justify their opinions. Such opinions are based on the territory, the dimensions of the land, the age of the building (adjusted to account for major renovations or additions), the quality of the construction and the components (materials) as well as the area of the buildings. As such, they provide information about the house, and fine-tuning a model on this data may produce a feature space that is more useful for home insurance pricing.

In our implementation, we use property tax data from the city of Québec in the province of Québec, Canada. We chose this dataset because it was readily available online and contained information for all 182,419 images within the property assessment dataset, with values corresponding to a market date of July 1, 2018. Table 1 summarizes the structured data available in this dataset. We only considered a subset of variables, including the number of floors, construction year, land value, building value, and total value. The reason behind this subset was the substantial amount of missing values in the other

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 1. *Variables and summary statistics from the property assessment dataset.*

Feature	Minimum	Average	Maximum	Standard deviation	% missing
Number of floors	1	1.55	33	0.86	3.51
Construction year	1604	1971	2018	30.15	8.52
Land value	1	238,387	175,781,000	1,226,238	0.15
Building value	1	435,890	652,421,000	5,047,336	0
Total value	1	673,896	828,202,000	6,123,145	0
Units	1	3.59	743	19.22	7.14
Non-residential units	1	2.89	416	8.92	92.81
Rental units	1	34.66	611	60.89	99.54
Front measure (m)	0.12	23.47	118,367	566.26	32.72
Area (m^2)	0.10	1,800.89	6,759,800	26,199.23	3.05

variables, which hindered precise modelling. We dropped any observations with missing values, which resulted in a training dataset size of 172,098. However, variables like the number of units and area were challenging to predict accurately due to the limitations of the data. For example, buildings with many units may be too tall to capture the entire structure in an image, while land area is difficult to predict using only a facade image and not a full property image. To limit the impact of outliers, we have applied several transformations to our dataset. Rather than using the construction year, we use the age of the building (with respect to the evaluation date of 2018). We have also capped the building age at 100 to ensure consistency and reduce the influence of outliers. Additionally, we have applied a logarithmic transformation to the land, building and total values to help normalize their distributions. However, we have removed five observations where the land, building or total value was listed as 1$, as these appeared to represent empty lots and may have skewed the analysis.

4.1.2. General structure of the representation models for computer vision

Models for computer vision are designed to analyse data such as images. These models typically use convolutional neural networks, a neural network suited to processing data with local information (Alzubaidi *et al.,* 2021). Image models based on deep learning have revolutionized computer vision by achieving state-of-the-art performance on many image-related tasks, such as image classification, image segmentation and object detection. These models are trained on large datasets of images and can automatically identify patterns relevant to a certain task. For an introduction to convolutional neural networks in actuarial science, see the supplementary materials of Blier-Wong *et al.* (2022) or Chapter 9 of Wüthrich and Merz (2023).

The general structure of convolutional neural networks (CNNs) has remained the same since the earliest CNN models (see, for instance, LeCun *et al.,* 1998). A typical image model contains two parts with trainable parameters; see Section 14.3 of Murphy (2022) for a review of popular architectures. The first is a set of convolutional layers which take as input an image and apply convolutional filters along the images. We call this the *Image model* in Figure 1. Such convolutional operations identify different patterns and shapes and, with enough depth, may construct a feature map that captures the information inside the image. If the image dataset contains diverse objects, then the feature map of the image will contain very general representations. We will refer to the first part of the image model as the convolutional part and the output of the first part as the feature map from the images. An intermediate part will flatten or unroll the feature map from the images (typically three-dimensional) into a single vector. The second part is a fully-connected set of neural network layers (usually only one) that will use the feature map from the images as input and a predictive task as output. For classification tasks, the output is one value for each category; for multitask regression, the output is one value for each regression task.

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

The fully-connected layer between the last hidden layer and the prediction has the same structure as a (generalized) linear model. Therefore, we can interpret the final hidden layer as a condensed representation that contains all of the useful information about the house image to predict the response variables in the related tasks. For this reason, we will consider the final hidden layer of the fully-connected set of layers to be the embedding associated with the house image. To recap, the fully-connected part of the representation framework has a set of layers devoted to simultaneously reducing the dimension of the feature space and capturing non-linear transformations and interactions between the feature maps from the images, which we call *Dimension reduction* in Figure 1. The output of this first part of fully-connected layers is the last hidden layer, which contains the representation of the house image, called *Image embedding* in Figure 1. The final fully-connected layer performs the role of a (generalized) linear model to predict the related tasks from the embeddings; we call this part the *Fully-connected predictor* in Figure 1.

4.1.3. Complete approach: fine-tuned image model

To make sure that the embeddings are useful for our eventual task of interest (predicting claim frequency/severity), we must "guide" our embeddings toward a feature space that is useful for insurance-related tasks, which implies that we must adapt the weights of the image model and fully-connected layers towards an attractive embedding space. To do this, we start with a pre-trained model that contains representations that are useful for general image classification. That is, the architecture of the set of convolutional layers remains the same. Then, we remove the original set of fully-connected layers and replace them with our own to control the size of the feature maps. The architecture of our fully-connected layers goes from the image model feature space size down to 128, then down to the embedding size and finally down to the size of the output task. The representation of the image will be the hidden features in the embedding layer, that is, the last hidden features before the predictions.

Within the complete approach, we allow the weights of the convolutional neural network, the fully-connected layers in the dimension reduction step and the fully-connected predictor to be trained on the related task. We apply the backpropagation algorithm to every trainable parameter in the first row of Figure 1. This model is the most flexible and takes the longest to train.

4.1.4. Limited approach: frozen image model

In the second example of unsupervised transfer learning, we use the same framework as the complete approach but allow less parameter flexibility. We keep the parameters of the convolutional neural network fixed and only train the new fully-connected layers. Within the framework outlined in Figure 1, this means that when constructing the model for the related task, one does not change the parameters from the image model (they remain the pre-trained parameters on ImageNet classification tasks). Still, one may train the parameters from the fully-connected layer in the dimension reduction part and the fully-connected predictor part. This means that feature maps from the images are kept the same as the original pre-trained image models.

In machine learning jargon, weights kept fixed are said to be frozen; hence, we call this model the frozen image representation. The limited approach is equivalent to extracting the flattened feature maps from each image in our dataset and training a fully-connected neural network according to the structure specified in the previous section. Therefore, for a fixed dataset, one may compute the feature map from the image model once, store this feature map and exclusively train the fully-connected neural network. From a training perspective, this makes training much more efficient since one does not have to pass the image in the image model (and apply the backpropagation algorithm to the weights of the image model) for every epoch. Within our implementation (that we will detail in Section 4.2), over 98% of the weights are in the image model.

We summarize the fine-tuned and frozen model in Figure 6. Within the smaller dotted box lies the fully-connected layers and the fully-connected predictor, whose weights are trainable with respect to the

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 2. *Summary of fine-tuned, frozen and PCA construction approaches.*

	Image model	Dimension reduction	Fully-connected predictor
Fine-tuned model	Flexible	Flexible	Flexible
Frozen model	Frozen	Flexible	Flexible
PCA with no fine-tuning	Frozen	PCA	NA

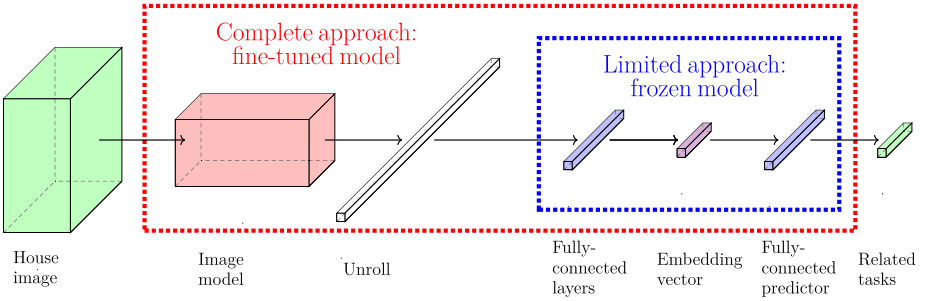

Figure 6. *Architecture for the complete and limited representation approaches.*

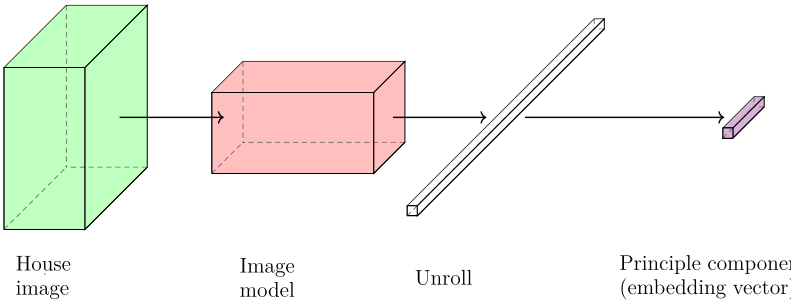

Figure 7. *Architecture for the basic PCA approach.*

related task. Similarly, within the larger dotted box, the weights from the image model are also trainable with respect to the same tasks.

4.1.5. Basic approach: PCA with no fine-tuning

The final approach to constructing embeddings is one where there is no fine-tuning in related tasks. To do this, we start with a pre-trained image model and use the feature map from the images. However, this representation is too large to include within the ratemaking model (512, 1024 or 2048 dimensions), so we apply principal component analysis to reduce the dimension to a more suitable embedding size. We present a diagram of the basic approach in Figure 7.

The basic approach, which we call the PCA approach, depends not on the related tasks (notice the absence of related tasks in Figure 7 compared to Figure 6) but on the representations learned on the ImageNet dataset. As such, the representations generated from this approach will capture general notions of shapes and categories but may not be directly appropriate for insurance tasks. However, it requires no fine-tuning tasks and almost no training time, making it a simple model to investigate whether there is useful information within images. We present a summary of the dimension reduction approaches in Table 2; this summary illustrates which parts of Figure 1 can be trained (flexible) and which maintains their pre-trained weights from ImageNet (frozen).

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

4.2. Experiments

4.2.1. Backbone image models

In this section, we will construct representations of image houses. For representation learning, it is the output of the first part that is of interest to us since this contains general information. We wish to use the embeddings within the representation pricing framework in Figure 1. We will consider some of the most popular image models, deep residual neural networks (ResNet, introduced by He *et al.,* 2016) and densely connected convolutional networks (DenseNet, introduced by Huang *et al*., 2017). For the ResNet models, one has pre-trained models for 18, 34, 50, 101 and 152 layers. The 152-layer model does not fit on our GPU, so we did not consider it. We also did not consider the 34-layer model to limit the length of the analysis. Pretrained DenseNet models are available as 121-, 169-, 201- and 264-layer models, but we only consider the 121-layer model since the larger ones did not fit on our GPU. All four models are pre-trained on the ImageNet dataset (Deng *et al*., 2009).

4.2.2. Training strategy

We have selected five related tasks to train the representation models: the construction year, the land value (log), the building value (log) and the total value (log), denoted, respectively by y_1, y_2, y_3 and y_4, and the number of floors. For the number of floors, we cap the values at three since the facade image for buildings with many floors will only show the first few floors. We consider the number of floors as a categorical variable (1, 2 or 3+ floors) and use a classification loss for this task; we denote these variables as y_5, y_6 and y_7. The number of observations for each class is 99,398, 59,915 and 12,785. The remaining tasks are treated as regression. The loss function is an equally weighted average of the mean squared error loss for the regression tasks and the cross-entropy loss for the classification task. We do not regularize the loss function. Therefore, the function to minimize is

$$\sum_{i=1}^{3}(y_i-\hat{y}_i)^2-\sum_{i=4}^{7}y_i\log\hat{y}_i,$$

where $\hat{y}_i$ are the predictions for y_i, for $i=1,\ldots,7$.

In our conversations with a few insurance companies, we know that typical ratemaking models for home insurance use between 30 and 100 rating variables. For that reason, we will consider embeddings of dimensions 8, 16 and 32 throughout this paper since we do not want to consider larger embedding dimensions such that the SVI features outnumber the traditional features.

To construct representations from unsupervised transfer learning, we follow the same training strategy for every model considered. We start with some pre-trained image models (ResNet and DenseNet) and get rid of the last layer (the one going from a feature space to a classification space). Instead, we add three fully-connected layers after the feature space, which we call the fine-tuning block. The first goes from the feature space size (depending on the model, 512, 1024 or 2048) to 128. Then, from 128 to the embedding size (8, 16 or 32; this acts as a hyperparameter), and the last from the embedding size to the output size (7). Between each layer, we use a LeakyReLU activation with a negative slope of 0.1, that is, $\max(0,x)-0.1\min(0,x)$. We initialize each fully-connected layer with Xavier initialization, proposed in Glorot and Bengio (2010), with the standard "gain" parameter of one. The bias for each fully-connected layer is initialized at zero.

To train the model, we set the batch size as the largest power of two such that the data and model fit on the GPU (GeForce RTX 20 with 8GB of RAM). We train the model for 25 epochs with an initial learning rate of 10^{-4}. Every five epochs, we reduce the learning rate by a factor of ten. We summarize the models in Table 3.

4.2.3. Experimental results

To answer our research question, we only require the intermediate representations to be related to insurance. For this reason, we do not need to (i) regularize the network, (ii) perform out-of-sample model

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 3. *Summary of parameters for ResNet and DenseNet models.*

	ResNet18	ResNet50	ResNet101	DenseNet121
Convolution weights	11,176,512	23,508,032	42,500,160	6,953,856
Image model feature space	512	2024	2024	1024
FC weights (8)	65,664	263,304	263,304	132,232
FC weights (16)	67,728	264,336	264,336	133,264
FC weights (32)	69,792	266,400	266,400	135,328
Batch size	128	32	32	32

Table 4. *Summary results on training set for Image models.*

		ResNet-18	ResNet-50	ResNet-101	DenseNet-121
Frozen image model	Loss (8)	1.261	1.281	1.252	1.184
	Loss (16)	1.267	1.199	1.197	1.162
	Loss (32)	1.172	1.161	1.143	1.125
	Training time	64 min	171 min	166 min	97 min
Fine-tuned image model	Loss (8)	0.078	0.075	0.051	0.098
	Loss (16)	0.055	0.069	0.051	0.082
	Loss (32)	0.055	0.069	0.050	0.077
	Training time	981 min	3779 min	5583 min	5076 min

validation, (iii) calibrate the predictions or (iv) find the best model or the best training strategy possible. Therefore, we only present results for in-sample training. Readers interested in information extraction from facade images should refer to Blanc (2022).

In Table 4, we summarize the loss values for the different models considered. Let us offer a few remarks on the performance of the information extraction models. First, the training time remained stable no matter the embedding size since they had approximately the same number of parameters (see Table 3). Training all models takes over 33 days on a GeForce RTX 20 with 8GB of RAM. For both the frozen image model and the fine-tuned image model, the best architecture is given by the ResNet-101 model with 32 embeddings, corresponding to the model with the most parameters. In general, increasing the size of the embedding layer leads to a lower loss, which makes sense since the models are more flexible. The only exception is for the frozen image model with ResNet-18, where the loss for embedding size 8 is lower than that for embedding size 16, which could be due to different local minima (training the same models with different initial weights should recover the correct ordering). For the fine-tuned image model, the difference in loss functions is very small, meaning that 16 embedding dimensions may be sufficient to learn the predictive tasks and that adding 16 more leads to redundant information. The DenseNet-121 model performed the best on the Frozen image model but worse on the fine-tuned image model. One possible explanation is that the DenseNet representations were more useful (which explains why they perform well on the image model) but that the higher number of parameters in the ResNet models enabled them to perform better when every parameter in the neural network was allowed to vary. Also, from Section 4.1, one observes that the percentage of variance captured by the first principal components of the feature space generated by the image models is larger for ResNet models compared with the DenseNet models, meaning that the feature space of DenseNet models are more linearly independent, meaning that DenseNet captures more nonlinear effects for the same embedding dimension.

We now provide prediction results for the ResNet-101 model with 32 embedding dimensions since it has the lowest loss value. In Figure 8, we present the confusion matrix for the number of stories in the frozen and fine-tuned image models. The root mean squared error (RMSE) for the regression tasks is in

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 5. *Root mean squared error of regression tasks for frozen and fine-tuned models with ResNet-101 and 32 embedding dimensions.*

	Age	Total value (log)	Building value (log)	Land value (log)
Frozen	25.91	0.5732	0.5422	0.7343
Fine-tuned	31.22	0.6580	0.3929	0.7703

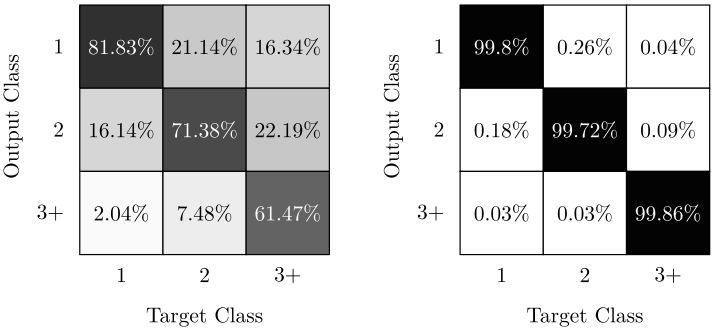

Figure 8. *Confusion matrices of the number of stories for frozen (left) and fine-tuned (right) models with ResNet-101 and 32 embedding dimensions on training set.*

Table 5. Recall that these values are for training datasets; we do not perform out-of-sample prediction since our primary goal is to construct useful representations. One notices from the confusion matrix that both models identify the correct class for most cases. Further, if a model misclassifies the number of floors, it is more likely to over/underestimate by one floor rather than by two floors. From Table 5, one observes that the RMSE for the regression tasks is quite high. One reason may be that age is a difficult category to predict since an old house could be renovated to look like new; hence, it is reasonable for a predictive model to predict a low age for a centenarian renovated house, yielding high RMSE values. Further, land value is highly dependent on the land size and the land location. While the image may contain information about the frontal measure of the land, it does not have access to information about the location. It may be this reason that land value is the worst regression task out of the three monetary variables. The building value is a proxy of the size and quality of the home, which is more useful for insurance contexts. For that variable, the fine-tuned model had a better predictive performance.

Let us finally look at the performance of PCA on the feature space from the images. In Figure 9, we present the ranked cumulative percentage of variance explained. While the curve of the cumulative percentage of variance explained has a similar shape for all models, the first principle components of the ResNet models capture more variance than the DenseNet model.

5. Actuarial application

Let us now summarize what we have accomplished and what is yet to come. We have collected an SVI dataset and a municipal evaluation dataset and cleaned out protected or temporary attributes from the images. Through three training strategies (PCA, frozen and fine-tuned), we have constructed embeddings of image houses. For the frozen and fine-tuned models, we trained these embeddings with the hope that they will be useful for insurance-related tasks such as predicting the number of stories in the house, the construction year and the building/land/total value of the property. As such, the representations will focus on the house and ignore other information which might not be related. The embeddings, therefore, act as proxies for the notion of the size of the house, the age or quality of the house and the value of the house. In this section, we will use embeddings as inputs to a ratemaking model. It is important to

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

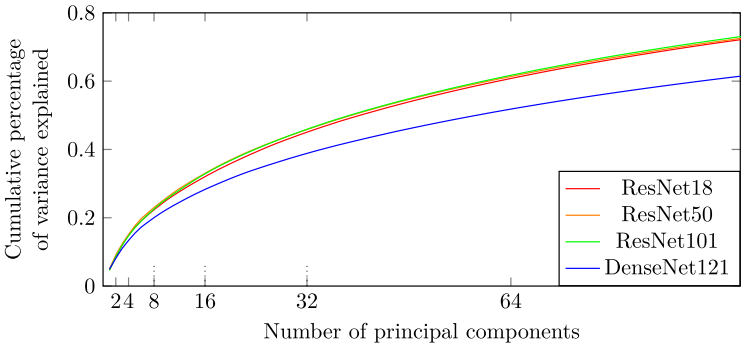

Figure 9. *Cumulative percentage of variance explained for the first principal components from the feature spaces.*

mention that for the entire actuarial application, the embeddings remain fixed and cannot be changed. Moreover, this marks our initial use of the insurance dataset, which implies that the claims data did not influence the construction of the embeddings. We perform the extrinsic evaluation of our embeddings on the insurance dataset. To do so, we follow the framework from Blier-Wong *et al*. (2021a, 2022) and train a GLM using only the embeddings.

5.1. Insurance data

We use a home insurance dataset from a large insurance company operating in Canada. We round the summary statistics such that they do not reveal customer information. The dataset contains information about individual perils, including `fire`, `other`, `water`, `wind`, `hail`, sewer backup (`SBU`) and `theft`. We do not know what type of coverages are included in the `other` peril. Finally, we create a category corresponding to the `total` perils. For each observation and each peril, we have the total frequency and total loss amounts. We consider the portion of the insurance portfolio located in Québec City, Québec, Canada, since this corresponds to the area where we have SVI. We have about 420,000 observations from 2009 to 2020, where the exposures for single observations vary from one day to one year (the mean and median exposures are, respectively, 0.44 and 0.42 years, meaning that most exposure periods are computed on a half-year basis). During this period, we possess SVI embeddings for around 40,000 distinct households.

Due to intellectual property constraints, we can only share relative data regarding the frequency and severity of various perils in our dataset. For each peril, the percentage of zeros is over 99%, indicating that we are dealing with infrequent events. In terms of frequency, `water` perils are the most common, occurring approximately two to three times more frequently than `theft`, `other` and `SBU` perils. The `wind` and `fire` perils are less common, with their frequency being 6–20 times lower than that of `water` perils. The least frequent peril is `hail`, which occurs 100 times less frequently than `water`. To analyse severity, we combine the `wind` and `hail` perils since there were too few observations to build a separate severity model for each. The median `total` severity is about 40% smaller than the average `total` severity, indicating positive skewness. The `theft`, `other` and `wind & hail` perils exhibit similar median severities. However, the median severities for `water`, `SBU` and `fire` perils are more than twice as high. The severity of fire risks is much more skewed towards higher values: even though the median severity for `fire` is similar to that of `water` and `SBU`, the average severity of `fire` is six times higher. This indicates that fire incidents may have many small damages but also very severe, for instance, in the case of the total destruction of the home. To limit the impact of outliers in the forthcoming gamma regression, we cap the severity at 100,000$ in our severity models. We split the dataset into two parts: one for training the ratemaking models and the other for evaluating the model on out-of-sample observations. The training/testing split is done on a 90%/10% basis.

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

To demonstrate the seamless integration of our ratemaking framework with existing actuarial pricing systems, we also include conventional actuarial variables in our ratemaking models. In our case, we use four variables: the client's age (`cage`), the number of years since the roofing was updated (`roof`), the age of the building (`bage`) and the limit amount of building coverage (`limit`). In the dataset provided, the client's age is in fifteen buckets of five-year intervals; therefore, we use dummy coding to construct features in the models. Building age was also a fine-tuning task (although this regression task had a high RMSE).

5.2. Frequency model

We now perform the extrinsic evaluation of the embedding models, that is, evaluate if the image embeddings accomplish the task of capturing useful information for insurance pricing. We start with a GLM for frequency modelling, using a Poisson response with the canonical link function

$$\ln(E[Y]) = \beta_0 + \ln\omega + \underbrace{\sum_{j=1}^{p} x_j\alpha_j}_{\text{traditional component}} + \underbrace{\sum_{k=1}^{\ell} \gamma_k\beta_k}_{\text{embedding component}}, \tag{5.1}$$

where ω is the exposure based on annual exposure base, p is the number of traditional features, ℓ is the embedding size for the image model, $x_j, j \in \{1,\dots,p\}$ are the traditional features and γ_k is the kth embedding dimension, $k \in \{1,\dots,\ell\}$. In (5.1), the embedding component (the scalar product between the embedding dimensions and their respective regression coefficients) corresponds to the contribution of the embeddings to the prediction. One can interpret the impact of the embeddings on the GLM prediction: the exponential of the embedding component has a multiplicative impact on the predicted frequency. If the exponential of the embedding component is larger (smaller) than one, then the impact of the SVI is to increase (decrease) the predicted frequency.

We compare each image model architecture (DenseNet121 and ResNet 18, 50 and 101) along with a baseline model with no embedding component. For each image model, we compare using an embedding size of 8, 16 and 32. In Tables 6, 7 and 8, we present the deviance on the test set for the PCA, frozen and fine-tuned method of constructing embeddings. Recall that a smaller deviance means a higher likelihood, meaning the model fit is better. In parenthesis, we include the number of GLM parameters from the embedding component significantly different than zero at a 0.05 significance level. If embeddings learned feature spaces that were linearly independent of the canonical function of the response variable, one would expect the number of significant GLM parameters from the embedding component at a 0.05 significance level to be 0.4, 0.8 and 1.6 for embedding sizes 8, 16 and 32. We highlight in bold the number of significant GLM parameters at a 0.05 significance level above the expected number of significant parameters under the assumption of independence. In every case, adding embeddings decreases the training deviance, which is unsurprising since there are more degrees of freedom in the models with embeddings. To limit the length of this paper, we do not show the training deviance.

What is immediately noticeable from Tables 6, 7 and 8 is that the embeddings greatly reduce the deviance for `SBU` and `water` claim frequency models. Increasing the embedding size provides better performance, meaning that the higher embedding dimensions do not learn redundant information (this point is even more valid for the frozen and fine-tuned models). For `SBU` and `water` perils, increasing the flexibility of the representation model leads to a better quality of representations to model claim frequency: for a fixed embedding size and model, going from PCA to frozen to fine-tuned models generally leads to better results, meaning that adapting the weights of the image models to focus on characteristics of the house leads to better representations of `SBU` and `water` damage claim frequency.

For the remaining perils, the results are more disputable. In most cases, we observe a slight increase or decrease in the deviance, so it is not clear if the embeddings capture a true effect within the SVI or if this is due to randomness. For instance, looking at the `theft` peril in the PCA approach, increasing

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 6. *Testing deviance for frequency prediction with fine-tuned models.*

Model	ℓ	Theft	Other	SBU	Water	Hail	Wind	Fire	Total
Baseline	0	1468.91	2028.72	2697.17	3831.07	163.47	884.08	479.96	8661.44
ResNet18	8	1468.06 (**1**)	2029.27 (0)	2670.60 (**3**)	3825.91 (**1**)	166.78 (0)	881.25 (**1**)	481.42 (**3**)	8665.47 (**2**)
	16	1468.12 (**8**)	2028.05 (**3**)	2667.97 (**7**)	3817.73 (**2**)	165.56 (0)	884.93 (0)	477.01 (0)	8657.23 (**7**)
	32	1460.68 (**2**)	2035.17 (**3**)	2662.17 (**9**)	3816.33 (**5**)	182.68 (**4**)	880.27 (0)	477.92 (**2**)	8656.09 (**2**)
ResNet50	8	1459.36 (**3**)	2030.22 (0)	2692.66 (**3**)	3825.50 (**1**)	163.87 (0)	885.85 (**1**)	477.11 (**3**)	8670.48 (**3**)
	16	1463.73 (**2**)	2028.17 (**2**)	2661.89 (**9**)	3820.33 (**7**)	170.42 (0)	884.61 (0)	475.02 (0)	8658.43 (**4**)
	32	1462.73 (**7**)	2029.40 (**3**)	2661.36 (**8**)	3816.36 (**2**)	177.95 (0)	878.15 (**2**)	481.98 (1)	8657.40 (**2**)
ResNet101	8	1460.39 (**4**)	2026.99 (**5**)	2671.30 (**5**)	3821.99 (0)	168.80 (0)	878.75 (**3**)	477.50 (**1**)	8659.38 (**3**)
	16	1464.28 (**7**)	2028.09 (0)	2663.49 (**6**)	3820.84 (**3**)	169.21 (0)	883.01 (**1**)	474.28 (0)	8659.49 (**6**)
	32	1468.98 (**3**)	2030.89 (**2**)	2665.01 (**8**)	3819.22 (**4**)	173.32 (**2**)	893.89 (1)	478.95 (1)	8664.47 (**10**)
DenseNet121	8	1473.56 (**5**)	2028.77 (**2**)	2672.94 (**5**)	3823.37 (**1**)	166.79 (**1**)	879.85 (**1**)	477.91 (0)	8664.57 (**2**)
	16	1467.33 (**2**)	2031.46 (**4**)	2665.68 (**6**)	3826.43 (0)	174.70 (0)	878.79 (**1**)	477.26 (0)	8661.09 (**4**)
	32	1477.16 (**6**)	2024.60 (**3**)	2668.62 (**7**)	3823.74 (**7**)	173.28 (0)	885.86 (0)	479.82 (0)	8660.95 (**5**)

Table 7. *Testing deviance for frequency prediction with frozen models.*

Model	ℓ	Theft	Other	SBU	Water	Hail	Wind	Fire	Total
Baseline	0	1468.91	2028.72	2697.17	3831.07	163.47	884.08	479.96	8661.44
ResNet18	8	1469.18 (**1**)	2031.90 (0)	2683.62 (**2**)	3828.06 (0)	161.78 (**1**)	882.82 (0)	472.73 (**1**)	8662.96 (**1**)
	16	1472.32 (**1**)	2027.60 (0)	2682.79 (**1**)	3824.34 (**4**)	162.67 (**1**)	885.53 (**1**)	475.82 (0)	8660.15 (**4**)
	32	1476.20 (1)	2032.81 (0)	2674.39 (**5**)	3830.33 (**3**)	175.35 (**2**)	883.74 (**2**)	466.54 (**3**)	8660.23 (**4**)
ResNet50	8	1469.59 (**1**)	2031.52 (0)	2681.21 (**4**)	3824.65 (**1**)	165.43 (0)	885.21 (0)	474.84 (0)	8661.91 (**2**)
	16	1472.47 (**3**)	2030.93 (**1**)	2677.99 (**2**)	3824.71 (**1**)	171.37 (0)	886.23 (**2**)	478.07 (**2**)	8659.54 (**1**)
	32	1477.54 (1)	2038.33 (**3**)	2679.71 (0)	3829.98 (1)	176.22 (**2**)	886.11 (**2**)	475.72 (0)	8665.00 (0)
ResNet101	8	1471.24 (**2**)	2031.64 (0)	2686.02 (**1**)	3823.83 (0)	169.96 (**3**)	883.29 (0)	477.17 (0)	8665.15 (0)
	16	1472.27 (**4**)	2031.03 (**2**)	2682.60 (**1**)	3823.30 (**1**)	169.02 (0)	881.50 (0)	475.97 (**1**)	8661.69 (**3**)
	32	1472.38 (0)	2028.80 (0)	2677.00 (**3**)	3828.52 (**5**)	177.78 (**3**)	890.98 (**2**)	480.66 (**4**)	8658.90 (1)
DenseNet121	8	1466.73 (**1**)	2030.89 (0)	2686.41 (**2**)	3826.29 (0)	164.15 (0)	883.90 (0)	477.74 (**1**)	8663.47 (**1**)
	16	1467.48 (**1**)	2034.43 (0)	2680.16 (**3**)	3825.57 (**3**)	160.94 (**2**)	887.59 (0)	480.37 (0)	8660.13 (**2**)
	32	1466.29 (**2**)	2038.45 (1)	2678.05 (**4**)	3829.30 (1)	164.56 (**2**)	889.04 (0)	484.92 (1)	8664.46 (**5**)

Table 8. *Testing deviance for frequency prediction with principal components.*

Model	ℓ	Theft	Other	SBU	Water	Hail	Wind	Fire	Total
Baseline	0	1468.91	2028.72	2697.17	3831.07	163.47	884.08	479.96	8661.44
ResNet18	8	1467.13 (0)	2029.41 (**1**)	2687.35 (**6**)	3824.90 (**2**)	162.36 (**1**)	882.65 (**2**)	470.51 (**1**)	8659.98 (**3**)
	16	1470.41 (0)	2028.73 (**1**)	2680.71 (**9**)	3823.31 (**2**)	163.68 (**2**)	881.86 (**3**)	468.21 (**1**)	8658.23 (**3**)
	32	1477.80 (**5**)	2030.89 (**3**)	2680.65 (**12**)	3827.10 (**2**)	172.65 (**4**)	879.11 (**4**)	466.83 (**2**)	8663.76 (**4**)
ResNet50	8	1465.87 (**1**)	2029.84 (0)	2684.10 (**5**)	3827.18 (0)	169.44 (**2**)	882.71 (**2**)	472.90 (**1**)	8659.97 (**2**)
	16	1467.17 (**2**)	2029.90 (0)	2678.30 (**8**)	3830.13 (**2**)	171.98 (**3**)	881.84 (**3**)	471.76 (**1**)	8663.14 (**4**)
	32	1471.53 (**4**)	2037.13 (**2**)	2677.54 (**12**)	3826.82 (**2**)	177.48 (**5**)	883.00 (**4**)	473.82 (**3**)	8662.24 (**6**)
ResNet101	8	1466.68 (0)	2028.41 (**1**)	2691.01 (**3**)	3828.88 (**2**)	163.32 (**1**)	883.10 (**1**)	472.99 (**1**)	8662.80 (0)
	16	1473.02 (**3**)	2029.71 (**1**)	2686.43 (**8**)	3828.70 (**3**)	171.36 (**3**)	884.29 (**2**)	475.40 (**1**)	8661.85 (**1**)
	32	1478.86 (**5**)	2033.07 (1)	2678.64 (**13**)	3824.33 (**5**)	168.79 (**4**)	880.54 (**3**)	473.26 (1)	8654.87 (**4**)
DenseNet121	8	1468.00 (**1**)	2031.26 (0)	2693.71 (**5**)	3827.15 (**1**)	164.41 (**1**)	883.80 (**1**)	479.15 (**2**)	8665.43 (**2**)
	16	1467.77 (**1**)	2037.79 (0)	2688.06 (**6**)	3824.19 (**1**)	162.47 (**1**)	884.06 (**1**)	477.76 (**1**)	8657.18 (**2**)
	32	1474.54 (**5**)	2041.15 (1)	2687.36 (**9**)	3826.71 (**3**)	165.85 (**1**)	881.89 (1)	476.55 (**2**)	8658.33 (**5**)

the embedding dimensions generally leads to an increase in deviance, meaning that the GLM may have overfit on the higher embedding dimensions. The best model for the `fire` peril is the PCA with the ResNet18 model. The fine-tuned models to predict the frequency of `fire` perils do not appear to increase or decrease the deviance systematically. Therefore, the auxiliary tasks we selected from the property tax data may not be related to the occurrence of fires. Most ResNet models from the fine-tuned approach for the `theft` and `wind` peril lead to decreased deviance compared to the baseline model. However, there is no clear-cut relationship between the embedding dimension and the deviance. This ambiguity makes understanding the role of the embeddings in the model harder, so someone might choose not to use them.

For every model, the `other` and `total` perils, there is no single model that is either better or worse than the baseline, implying that the embeddings are not useful within the GLM model. We find this surprising: the embeddings unquestionably improved the `SBU` and `water` perils, which are the most common ones in the dataset. However, it seems that combining the remaining perils into one regression task eliminates that improvement. We, therefore, stress that embeddings have different effects on each peril because the embeddings are not useful for the "catch-all" perils `other` and `total`.

The only peril for which embeddings usually increase the deviance is `hail`. This peril occurs when a hailstorm causes damage to a house, say, if a large hailstone falls on a roof or shatters a window. The increase in deviance may be due to a lack of causal effect (there may be no link between the picture of a house and the probability of filing a claim due to hail). Another explanation is that the hail peril is the rarest in the dataset: hailstorms are infrequent in Québec city (see, for instance, Etkin, 2018 for some frequency statistics). When they do occur, the size of hailstones is typically too small to cause a loss larger than the deductible amount.

When looking at the number of significant parameters at the 0.05 level for the embedding component, one observes that this number, for most models, is higher than the expected number under the independence assumption. For instance, for the fine-tuned models, almost all of the models for `theft`, `SBU`, `water` and `total` perils had a high number of significant GLM parameters for the embedding component, even if we could only conclude that the `SBU` and the `water` claims had clear improvements in the models. Further, for the `SBU` peril, one usually has more significant parameters for the PCA approach compared with the fine-tuned approach, even though the fine-tuned yields smaller deviance values. Therefore, the statistical significance of the embedding parameters may not be a good extrinsic evaluation measure for embeddings. However, it would be cause for alarm if a model with embeddings yielded a much smaller deviance value if none of the parameters associated with the embedding components were statistically significant.

5.3. Effect of including variables in both relevant tasks and traditional variables

Let us next look at the perils for which the SVI embeddings provided the best improvements and their effect on the age of the building. Recall that the age of the building was a fine-tuning task from the tax assessments; hence the image representations with the frozen and fine-tuned approaches capture representations of the age (even if the RMSE of these models was high). Therefore, the model containing embeddings constructed with the frozen and fine-tuned approaches uses both the `bage` variable and a proxy to the `bage` variable through the unsupervised transfer learning. Note, however, that the variable in the insurance dataset is dynamic (age changes every year since we have observations between 2009 and 2020), while the proxy in the embeddings is static (since it depends on the age of the building at the tax assessment date). In Table 9, we present the p-value associated with the variable `bage` for the baseline model and the three embedding construction approaches. Note that the `bage` variable is significant in both models (assuming a 0.05 significance level). We consider only ResNet18 and ResNet101 models with an embedding dimension of 32 to limit space, but one obtains similar results for other embedding sizes and image model architectures. Further, we only consider the `SBU` and `water` perils since they are the ones for which SVI embeddings improved the deviance for every set of embedding. For the `SBU`

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 9. *Comparison of p-values for the variable* `bage` *with and without embeddings.*

		PCA		frozen		fine-tuned	
	Baseline	ResNet18	ResNet32	ResNet18	ResNet32	ResNet18	ResNet32
`SBU`	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	0.0005912	0.000018
`water`	0.006181	0.312924	0.280731	0.390484	0.943965	0.445789	0.704342

peril, including the embeddings, does not deem the `bage` variable insignificant since they remain under the 0.05 threshold for every embedding construction method, image model architecture and embedding size. For the `water` peril, including any embedding makes the `bage` variable insignificant, meaning that the embeddings may be a better proxy to any risk-generating process than the age of the building. We remark that the *p*-values for the frozen and fine-tuned model are higher than for the PCA models, but one should not confuse this with the notion of having "less significance" but that no effect was observed. From this experiment, we cannot conclude if the image model or the unsupervised transfer learning component of the representation learning framework yielded the `bage` insignificant for the `water` peril.

Another diagnostic that will help us interpret the results of the regression is the variance inflation factors (VIF), more specifically, their generalized version introduced in Fox and Monette (1992) since the `cage` attribute contains 14 degrees of freedom. In Table 10, we present the VIF for the baseline model and different embedding construction methods for the ResNet18 model with eight embedding dimensions. We also present results for models trained with the embedding component only, that is, without the traditional component.

Recall that a VIF over ten is considered problematic and that a cut-off of 5 is often recommended. The VIFs in the baseline model are all around one, indicating low collinearity. The VIF for each variable in the traditional component usually increases when adding the embedding component and vice versa. Let us now use VIFs to examine the effect of including a variable in the regression model and the related tasks. The VIF for `bage` increases as the embedding model increases. For the most flexible embedding construction method, the VIF is about four times as large as the baseline, indicating moderate collinearity and hinting that the `bage` variable may become duplicated within the embedding dimensions. Note that the VIF for the embedding components is very high: we will examine the reasons and propose a solution in the following subsection.

5.4. Impact of correlated embeddings

One aspect to consider for applications of embeddings is the correlation of embedding dimensions. If the correlation is too high, there may be collinearity in the features, which could cause the variance of regression coefficients to be inflated (as observed in Table 10). Note that, by construction, PCA embeddings are orthogonal to each other (hence linearly uncorrelated), so we must only diagnose the correlation of the embeddings generated by unsupervised transfer learning. In Figure 10, we present the correlation matrix for the ResNet18 embeddings with eight embedding dimensions for the frozen and fine-tuned approaches of embedding construction. One observes a linear correlation (between -0.75 and 0.91) between the embedding dimensions. One way to remove this linear correlation is to use all of the principal components of the embeddings instead of the embeddings themselves. While this approach does not make much sense in typical GLM modelling (since one would lose the ability to interpret the model and perform variable selection), doing so on image embeddings does not hurt the model since the image embeddings are already uninterpretable.

In Table 11, we present the VIFs for the GLM trained on decorrelated embedding dimensions. One observes that the VIF for the traditional component stays the same if one uses the frozen/fined-tuned embeddings or their principal components. The VIFs in Table 10 for the frozen and fine-tuned

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 10. *Variance inflation factors for different embedding construction approaches.*

	offset	`cage`	`roof`	`bage`	`limit`	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8
Baseline	1.006	1.080	1.018	1.116	1.043								
PCA	1.000					1.274	1.172	1.031	1.019	1.034	1.155	1.003	1.043
	1.007	1.098	1.018	1.273	1.097	1.302	1.263	1.045	1.057	1.055	1.158	1.008	1.050
Frozen	1.000					7.215	38.938	14.329	13.733	26.833	54.160	5.109	16.064
	1.007	1.101	1.017	1.610	1.153	7.323	40.793	14.467	13.846	27.622	55.760	5.996	16.296
Fine-tuned	1.000					57.250	23.677	60.947	6.662	7.675	104.657	55.757	34.132
	1.011	1.120	1.017	4.410	2.084	58.914	31.823	63.224	6.887	8.100	106.449	62.353	52.510

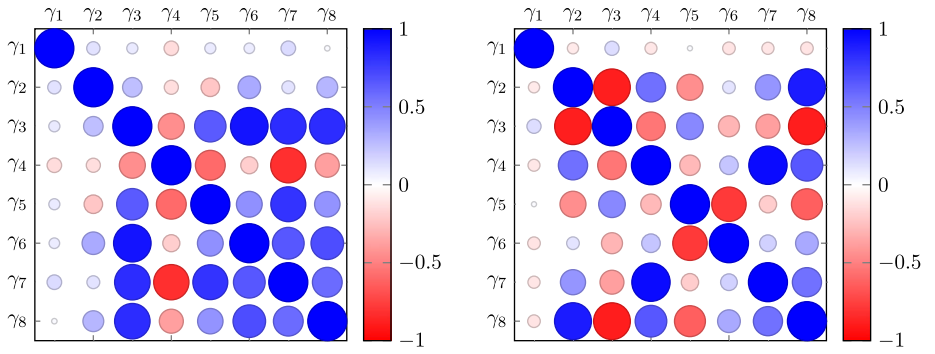

Figure 10. *Correlation matrix of embedding dimensions for ResNet-18 with eight embeddings for frozen (left) and fine-tuned (right) approaches.*

embedding models are high, always over the cut-off of 5 and sometimes reaching over 100. However, when using the principal components of the frozen and fine-tuned embeddings, the VIFs are no longer considered high. It follows that the variance of the predictors is lower, implying that more variables become statistically significant. Note that this step does not impact the predictive variance but provides a more useful way to diagnose the statistical significance of embedding parameters and the effect of the collinearity of the embeddings on the GLM.

5.5. Severity model

We next extrinsically evaluate the quality of severity models. In this case, we use a gamma response with the canonical link function:

$$E[Y]^{-1} = \beta_0 + \sum_{j=1}^{p} x_j \alpha_j + \sum_{k=1}^{\ell} \gamma_k \beta_k. \tag{5.2}$$

In (5.2), the embedding component (the scalar product between the embedding dimensions and their respective regression coefficients) corresponds to the contribution of the embeddings to the score function $E[Y]^{-1}$. In Tables 12, 13 and 14, we present deviant results on the test dataset with the PCA, frozen and fine-tuned approaches, respectively. Recall that since the number of `wind` and `hail` perils was too small, we trained a combined model for these two perils. Further, some GLM models did not converge for the `fire` peril with the PCA approach to construct embeddings; we denote these by NA in the deviance result tables.

Overall, the results for the severity models are much less impressive than those for the frequency models. The only exception is for the `SBU` peril, where one mostly observes a slight reduction in deviance and where unsupervised transfer learning improves the results compared with PCA. In general, one could conclude that the embeddings do not contribute much to the baseline severity model. One reason may be the limited dataset size (some perils have under 100 observations) or the embeddings do not capture useful features for severity modelling.

6. Discussion

We have proposed a simple model to use images within a ratemaking model. This approach does not drastically increase the number of parameters within the predictive model. We find that images improve the predictive ability of ratemaking models, meaning that there are observable characteristics within images that affect the risk of insurance contracts.

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 11. *Variance inflation factors after decorrelating embeddings.*

	offset	`cage`	`roof`	`bage`	`limit`	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8
Frozen + PCA	1.000					1.848	2.186	1.080	1.345	2.400	1.835	1.185	1.320
	1.007	1.101	1.017	1.610	1.153	1.912	2.344	1.129	1.708	2.441	1.828	1.186	1.312
Fine-tuned + PCA	1.000					1.717	1.963	2.493	2.276	1.280	1.448	3.096	1.302
	1.011	1.120	1.017	4.410	2.084	1.943	2.109	5.053	3.094	2.448	1.535	3.219	1.313

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 12. *Testing deviance for severity prediction with fine-tuned models.*

Model	ℓ	Theft	Other	SBU	Water	Wind & Hail	Fire	Total
Baseline	0	95.24	225.71	298.61	521.48	46.74	47.28	1220.89
ResNet18	8	106.03 (0)	231.51 (0)	296.43 (0)	520.11 (0)	46.97 (0)	75.23 (**1**)	1225.91 (0)
	16	107.42 (**2**)	230.90 (**2**)	297.50 (**2**)	520.24 (0)	49.71 (**1**)	194.07 (**1**)	1216.52 (0)
	32	105.78 (**2**)	240.60 (**2**)	296.68 (**2**)	523.33 (0)	60.16 (**5**)	120.04 (**2**)	1226.94 (**4**)
ResNet50	8	107.29 (0)	226.51 (0)	293.31 (0)	521.60 (0)	47.01 (0)	56.13 (**2**)	1223.54 (0)
	16	103.91 (**1**)	223.79 (**3**)	295.71 (**1**)	517.79 (**1**)	50.13 (0)	67.36 (**1**)	1230.12 (**5**)
	32	104.82 (1)	227.82 (**8**)	294.36 (1)	534.31 (**2**)	51.13 (0)	64.48 (0)	1224.14 (**2**)
ResNet101	8	99.50 (0)	225.43 (0)	293.50 (0)	519.70 (0)	48.59 (0)	102.99 (**1**)	1227.88 (0)
	16	98.92 (**1**)	228.82 (**1**)	291.73 (**1**)	523.08 (**2**)	48.75 (**2**)	77.04 (**4**)	1227.89 (**1**)
	32	108.00 (1)	222.63 (0)	293.31 (**6**)	526.59 (**2**)	45.69 (1)	307.10 (**5**)	1240.69 (1)
DenseNet121	8	98.31 (**2**)	230.58 (**1**)	293.14 (0)	518.43 (0)	47.85 (0)	55.11 (0)	1217.55 (**2**)
	16	102.45 (**1**)	228.88 (0)	291.90 (**2**)	519.27 (0)	46.95 (**1**)	63.48 (**1**)	1228.68 (**2**)
	32	92.31 (0)	244.53 (**8**)	293.50 (**3**)	525.44 (**2**)	53.29 (0)	112.33 (**2**)	1228.05 (0)

Table 13. *Testing deviance for severity prediction with frozen models.*

Model	ℓ	Theft	Other	SBU	Water	Wind & Hail	Fire	Total
Baseline	0	95.24	225.71	298.61	521.48	46.74	47.28	1220.89
ResNet18	8	97.34 (0)	221.27 (**1**)	292.41 (**1**)	521.37 (0)	46.10 (0)	45.38 (**6**)	1220.12 (**1**)
	16	102.85 (**1**)	223.97 (**2**)	288.03 (**4**)	519.49 (0)	49.28 (0)	40.67 (**1**)	1216.13 (**1**)
	32	94.56 (**2**)	225.00 (0)	295.39 (**2**)	519.08 (0)	60.51 (9)	69.46 (1)	1215.95 (0)
ResNet50	8	92.96 (0)	233.08 (**1**)	297.06 (**1**)	516.06 (0)	50.16 (0)	53.33 (0)	1221.52 (**2**)
	16	96.56 (0)	229.65 (0)	297.46 (**2**)	516.49 (0)	52.40 (0)	61.37 (**4**)	1221.67 (**1**)
	32	105.37 (**3**)	242.79 (**5**)	296.65 (**2**)	525.25 (1)	54.33 (**3**)	87.14 (**4**)	1221.94 (0)
ResNet101	8	95.73 (0)	230.53 (0)	298.55 (0)	519.11 (0)	53.48 (0)	50.11 (0)	1218.42 (0)
	16	97.68 (0)	224.30 (**1**)	299.70 (**2**)	514.68 (0)	53.72 (**2**)	42.55 (0)	1222.20 (**1**)
	32	96.85 (**4**)	238.12 (**6**)	299.13 (0)	520.12 (**2**)	55.85 (**2**)	43.36 (**2**)	1214.32 (1)
DenseNet121	8	96.85 (0)	232.33 (**2**)	298.37 (**1**)	518.93 (0)	49.26 (0)	49.92 (0)	1226.32 (0)
	16	97.15 (0)	238.72 (**3**)	297.81 (**1**)	520.40 (0)	51.92 (0)	64.97 (**3**)	1220.43 (**1**)
	32	101.36 (1)	234.20 (**3**)	295.07 (**2**)	510.46 (**2**)	61.74 (0)	58.86 (**4**)	1229.52 (**2**)

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Table 14. *Testing deviance for severity prediction with principal components.*

Model	ℓ	Theft	Other	SBU	Water	Wind & Hail	Fire	Total
Baseline	0	95.24	225.71	298.61	521.48	46.74	47.28	1220.89
ResNet18	8	90.89 (0)	234.88 (**1**)	295.44 (**4**)	521.16 (0)	46.86 (**1**)	53.01 (**2**)	1227.32 (**2**)
	16	95.16 (0)	226.80 (**1**)	295.27 (**5**)	522.05 (**2**)	51.90 (**2**)	55.51 (**2**)	1227.56 (**3**)
	32	103.94 (1)	229.50 (**2**)	292.98 (**5**)	529.70 (**3**)	59.47 (**3**)	NA	1236.52 (**6**)
ResNet50	8	90.07 (0)	234.23 (**1**)	299.18 (**4**)	523.22 (0)	46.38 (**1**)	44.54 (**2**)	1224.86 (**2**)
	16	89.60 (**1**)	235.49 (**1**)	297.73 (**5**)	521.87 (**1**)	52.65 (**2**)	81.20 (**5**)	1224.32 (**3**)
	32	96.21 (**4**)	233.98 (**2**)	298.03 (**5**)	530.90 (**2**)	67.74 (**5**)	75.43 (**4**)	1224.57 (**6**)
ResNet101	8	92.97 (**1**)	226.25 (0)	300.94 (**3**)	525.15 (**1**)	46.60 (**1**)	57.26 (**1**)	1230.82 (**2**)
	16	94.56 (**1**)	227.19 (0)	300.49 (**3**)	522.54 (0)	56.20 (**2**)	NA	1229.13 (**2**)
	32	97.35 (1)	224.73 (**3**)	300.31 (**5**)	521.75 (1)	69.21 (**5**)	76.24 (**2**)	1227.08 (**4**)
DenseNet121	8	94.61 (**3**)	231.82 (**1**)	298.49 (**2**)	520.42 (0)	47.21 (**1**)	50.15 (**1**)	1229.60 (**2**)
	16	98.96 (**2**)	231.40 (**2**)	298.98 (**2**)	518.73 (**1**)	51.02 (**2**)	NA	1233.66 (**1**)
	32	101.26 (**4**)	224.64 (**4**)	306.69 (**2**)	518.62 (1)	58.70 (**5**)	68.30 (**6**)	1233.37 (**2**)

We find statistically significant evidence for a relationship between image data and claim counts for certain perils. This is not the case for claim severity models. Our approach relies on embeddings, so we cannot conclude that there is a causal relationship between the image data and the claims count data. Our work, however, finds predictive power in the images, which means that there could be some phenomena in the images which have a causal impact on losses. Future work could investigate which parts of the image have causal impacts on premiums, such that insurance companies could start collecting this information to use within pricing models.

We also note that we considered other sources of images to attempt to answer our research question. First, we tried to use images from real estate websites. The advantages of such images are that they are of high quality, the pictures of facades are well framed in the image, and we have access to much-structured information that one could use for fine-tuning the image models to related tasks. A disadvantage of this approach is that a limited number of houses have a real-estate listing. If an insurance company attempts to use images to provide a quote to a potential customer, the image of that customer's home could not be available from one of the real-estate websites. Therefore, we needed to use a data source available for most of the potential customers in a region. Then, we considered using aerial imagery (for instance, Google Satellite). Note that most of these images for residential areas are not taken by satellites but by planes flying at low altitudes with high-resolution cameras. The advantage of this approach is that aerial imagery is available for most cities on Google Satellite. Otherwise, an insurance company can access a data provider for higher-quality data imagery. See, for instance, Liu *et al.* (2023) for a review of CNNs to aerial imagery, including applications such as object detection, classification and semantic segmentation. We decided not to use aerial imagery since we have not found useful, structured information related to insurance such that we may fine-tune the image representations to insurance-related tasks. However, we believe that aerial imagery will become an essential tool for risk assessment and insurance pricing since one could extract useful information such as land area, house area, presence or not of additions such as garages or pools or the presence or not of objects that could cause claims such as trees or electric poles.

Our goal in this research was to provide empirical evidence of useful information in SVI for insurance pricing. Therefore, this exploratory work serves as a proof-of-concept for this ratemaking framework. For this reason, we construct the embeddings using data freely available online for anybody to access. With the demonstrated usefulness of image data for insurance ratemaking, insurance companies might consider investing additional time and money in collecting and using this data source. For instance, one could attempt to obtain better quality images by building their dataset of SVI by taking individual pictures of houses that are all high quality and centered on the house of interest. Also, we fine-tuned our image models on tax assessment datasets, but insurance companies already have access to features of interest for homes they insure since they collect data about the house during the quoting process. Therefore, one could construct representations of images using more insurance-related tasks.

As a secondary objective of our project, we showed that one could use SVI to extract useful information about houses using online data automatically. Since our fine-tuning data were limited, we could only extract information like the construction year, the number of stories and the land/building/total value. However, insurance companies have access to internal data collected during the quoting process, such as the roofing material, the facade material and the presence or not of garages, sheds or pools. Therefore, another use of our framework is automatically extracting features from online images. In that case, insurance companies could attempt to predict some characteristics of the insured house without asking the customer to fill out the information manually. For instance, if the image model is confident that the facade type is made of a specific material, that field in the quoting form could be pre-filled with that material, improving the customer experience.

The representation learning framework is beneficial since it lets us empirically verify that SVI contains valuable information to predict the claim frequency of home insurance contracts. However, since image models include so many parameters (even the smallest model we consider has over 6 million parameters in the convolutional weights alone), they are prone to overfit on the predictive task. For

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

this reason, we train the model on related tasks such that the associated representations are adapted to insurance-related tasks and avoid overfitting on frequency or severity prediction.

The objective of these experiments was not necessarily to encourage insurance companies to use images within their ratemaking process but to show empirical evidence that useful information within images leads to better actuarial fairness. For example, we have shown that SVI is useful for predicting claim frequency for specific perils and works especially well for sewer backup claims. In future work, one should seek to identify the most informative parts of the image, providing insights into the underlying mechanisms that drive risk in home insurance. In that case, one would retain the increased actuarial fairness but could also interpret and communicate these results convincingly to management, regulators and customers, opening the door to more informed decision-making.

Acknowledgements. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (CRDPJ 515901-17, Blier-Wong: 559169, Marceau: 05605). We thank Intact Financial Corporation for the data, support and comments from Eliane Belisle, Frédérique Paquet, and Étienne Girard-Groulx. We thank Cyril Blanc for his help with the dataset and Ronald Richman for his insightful comments. We thank the anonymous referees for fruitful comments that have improved the quality of this paper.

Competing interests. The author(s) declare none.

References

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L. (2021) Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, **8**(1), 53.

Avanzi, B., Taylor, G., Wang, M. and Wong, B. (2023) Machine learning with high-cardinality categorical features in actuarial applications. arXiv preprint arXiv:2301.12710.

Bengio, Y., Courville, A. and Vincent, P. (2014) Representation Learning: A Review and New Perspectives. arXiv:1206.5538 [cs].

Biffis, E. and Chavez, E. (2017) Satellite data and machine learning for weather risk management and food security: Satellite data and machine learning for weather risk management and food security. *Risk Analysis*, **37**(8), 1508–1521.

Biljecki, F. and Ito, K. (2021) Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, **215**, 104217.

Blanc, C. (2022) *Caractérisation automatique d'immeuble depuis une image de façade*. Master's Thesis, Université Laval.

Blier-Wong, C., Baillargeon, J.-T., Cossette, H., Lamontagne, L. and Marceau, E. (2021a) Rethinking representations in P&C actuarial science with deep neural networks. arXiv:2102.05784 [stat].

Blier-Wong, C., Cossette, H., Lamontagne, L. and Marceau, E. (2021b) Machine learning in P&C insurance: A review for pricing and reserving. *Risks*, **9**(1), 4.

Blier-Wong, C., Cossette, H., Lamontagne, L. and Marceau, E. (2022) Geographic ratemaking with spatial embeddings. *ASTIN Bulletin*, **52**(1), 1–31.

Brock Porth, C., Porth, L., Zhu, W., Boyd, M., Tan, K.S. and Liu, K. (2020) Remote sensing applications for insurance: A predictive model for pasture yield in the presence of systemic weather. *North American Actuarial Journal*, **24**(2), 333–354.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. and Amodei, D. (2020) Language models are few-shot learners. Advances in Neural Information Processing Systems 33, pp. 1877–1901.

Chen, F.-C., Subedi, A., Jahanshahi, M.R., Johnson, D.R. and Delp, E.J. (2022) Deep learning– based building attribute estimation from Google street view images for flood risk assessment using feature fusion and task relation encoding. *Journal of Computing in Civil Engineering*, **36**(6), 04022031.

Delong, Ł. and Kozak, A. (2023) The use of autoencoders for training neural networks with mixed categorical and numerical features. *ASTIN Bulletin: The Journal of the IAA*, **53**(2), 1–20.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009) ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. IEEE.

Doshi, S., Gupta, A., Gupta, J., Hariya, N. and Pavate, A. (2023) Vehicle damage analysis using computer vision: Survey. *2023 International Conference on Communication System, Computing and IT Applications (CSCITA)*, pp. 132–135.

Embrechts, P. and Wüthrich, M.V. (2022) Recent challenges in actuarial science. *Annual Review of Statistics and Its Application*, **9**(1), annurev–statistics–040120–030244.

Etkin, D. (2018) *Hail climatology for Canada: An update*. Institute for Catastrophic Loss Reduction.

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press

Fang, F., Yu, Y., Li, S., Zuo, Z., Liu, Y., Wan, B. and Luo, Z. (2021) Synthesizing location semantics from street view images to improve urban land-use classification. *International Journal of Geographical Information Science*, **35**(9), 1802–1825.
Fox, J. and Monette, G. (1992) Generalized collinearity diagnostics. *Journal of the American Statistical Association*, **87**(417), 178–183.
Gao, G. and Wüthrich, M. (2019) Convolutional neural network classification of telematics car driving data. *Risks*, **7**(1), 6.
Glorot, X. and Bengio, Y. (2010) Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings.
He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708.
LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
Lee, G.Y., Manski, S. and Maiti, T. (2020) Actuarial applications of word embedding models. *ASTIN Bulletin*, **50**(1), 1–24.
Liu, X., Ghazali, K.H., Han, F. and Mohamed, I.I. (2023) Review of CNN in aerial image processing. *The Imaging Science Journal*, **71**(1), 1–13.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J. (2013) Distributed representations of words and phrases and their compositionality. *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems*, p. 9.
Murphy, K.P. (2022) *Probabilistic Machine Learning: An Introduction*. Cambridge, MA: The MIT Press.
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I. (2019) Language models are unsupervised multitask learners. *OpenAI Blog*, **1**(8), 9.
Richman, R. (2020a) AI in actuarial science – a review of recent advances – part 1. *Annals of Actuarial Science*, **15**(2), 1–23.
Richman, R. (2020b) AI in actuarial science – a review of recent advances – part 2. *Annals of Actuarial Science*, **15**(2), 1–29.
Shi, P. and Shi, K. (2022) Non-life insurance risk classification using categorical embedding. *North American Actuarial Journal*, **27**(3), 1–23.
Tian, Y., Wang, Y., Krishnan, D., Tenenbaum, J.B. and Isola, P. (2020) Rethinking few-shot image classification: A good embedding is all you need? *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, pp. 266–282. Springer.
Turian, J., Ratinov, L. and Bengio, Y. (2010) Word representations: A simple and general method for semi-supervised learning. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 384–394.
Wüthrich, M.V. (2017) Covariate selection from telematics car driving data. *European Actuarial Journal*, **7**(1), 89–108.
Wüthrich, M.V. and Merz, M. (2023) *Statistical Foundations of Actuarial Learning and Its Applications*. *Springer Actuarial*. Cham: Springer International Publishing.
Wüthrich, M.V. and Ziegel, J. (2023) Isotonic recalibration under a low signal-to-noise ratio. arXiv preprint arXiv:2301.02692.
Xu, S., Zhang, C. and Hong, D. (2022) BERT-based NLP techniques for classification and severity modeling in basic warranty data study. *Insurance: Mathematics and Economics*, **107**, 57–67.
Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A. and Torralba, A. (2017) Scene parsing through ADE20K dataset. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 633–641.
Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A. and Torralba, A. (2019) Semantic understanding of scenes through the ADE20K dataset. *International Journal of Computer Vision*, **127**, 302–321.
Zhu, R. and Wüthrich, M.V. (2021) Clustering driving styles via image processing. *Annals of Actuarial Science*, **15**(2), 276–290.
Zhu, W. (2023) A deep factor model for crop yield forecasting and insurance ratemaking. *North American Actuarial Journal*, **28**(1), 1–16.

https://doi.org/10.1017/asb.2024.9 Published online by Cambridge University Press