Check for updates

Geographical Analysis

ORIGINAL ARTICLE OPEN ACCESS

Capturing Non-Linear Neighborhood Structure: An Autoencoder Approach to Census-Based Dimensionality Reduction for Residential Differentiation

Alex Singleton[1] | Owen Goodwin[1] | Stef De Sabbata[2,3]

[1]Department of Geography and Planning, University of Liverpool, Liverpool, UK | [2]School of Geography, Geology and the Environment, University of Leicester, Leicester, UK | [3]Geolytix, London, UK

Correspondence: Alex Singleton (alex.singleton@liverpool.ac.uk)

Received: 3 March 2026 | **Revised:** 7 September 2026 | **Accepted:** 14 September 2026

ABSTRACT

Dimensionality reduction is fundamental to applied spatial data analysis, condensing high-dimensional indicators into parsimonious representations for mapping and modeling. Principal component analysis (PCA) remains a dominant approach, yet its linearity assumptions constrain its capacity to capture the complex, non-linear dependencies characteristic of spatial socio-economic data. We evaluate the effectiveness of autoencoders as a flexible, parametric alternative for constructing composite measures of neighborhood structure. Through encoder-decoder architectures, AEs learn non-linear representations that better reflect underlying data complexity, enable projection of new observations without refitting, and provide reconstruction-based diagnostics that enhance transparency. We demonstrate the approach using small-area data from the 2021 Census of England and Wales, evaluating how effectively AEs summarize spatial variability and recover composite dimensions of residential differentiation compared with PCA.

1 | Evolving Approaches to the Measurement of Composite Neighborhood Structure

A composite measure of neighborhood structure is a numerical representation derived from underlying patterns in multidimensional spatial data, capturing key geographic characteristics of neighborhoods, such as demographic composition, socio-economic status, or built environment attributes. Unlike a univariate measure, a composite can capture a broad range of characteristics, including interactions, and represent more abstract latent concepts that are not directly measurable such as "deprivation." Within a contemporary geographic context, composite measures typically emerge from the application of dimensionality reduction techniques to identify and quantify otherwise hidden spatial relationships or structures. The development of composite indicators is well established within the academic literature, and there is a lineage of different methodological approaches (Harris et al. 2005; Singleton and Spielman 2014, 2026; Batey 2022). Social Area Analysis was developed by Shevky and Williams (1949) and had a particularly strong influence in early work from the 1950s and 1960s. Although this drew criticism regarding theoretical robustness and generalizability across contexts (Ericksen 1949; Hawley and Duncan 1957; Bassett and Short 1980), Social Area Analysis established foundational methodology for the integration of multivariate census data into a spatial analytical framework (Shevky and Bell 1955; Rees 1971; Timms 1971). Factorial Ecology emerged later during the 1970s, and extended Social Area Analysis through the integration of factor analysis and a relaxation of input measure specifications (Berry and Kasarda 1977; Hale and Austin 1997). This was extensively applied through the 1980s and continues in limited contemporary studies (Al-Shawamreh and

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Geographical Analysis* published by Wiley Periodicals LLC on behalf of The Ohio State University.

Farhan 2018; Salvati et al. 2018; Hirokane and Amemiya 2023; Montosa Muñoz 2023). However, Factorial Ecology has also faced critique on issues of model interpretability, generalizability, and reproducibility (Johnston 1977; Harris et al. 2005). Furthermore, given that spatial datasets are typically characterized by complex non-linear relationships, scholarship has increasingly emphasized the limitations of linear techniques in capturing spatial patterns and dependencies (Hughes and Carey 1972; Hunter 1972; Da Vies 1978).

1.1 | The Case for Non-Linear Representations

The composite measures of neighborhood structure discussed in the preceding section have prevalently been underpinned by linear dimensionality reduction techniques, particularly Principal Component Analysis (PCA) and related factor analysis methods. PCA operates by computing a set of new orthogonal variables, known as principal components, that are linear combinations of the original variables (Jolliffe 2002). Each component is constructed to maximize variance, and the components are ordered such that the first captures the largest share of total variance, the second the next largest, and so forth. Dimensionality reduction is then achieved by selecting the first n components that together explain the majority of variance, allowing a high-dimensional dataset to be represented by a smaller number of composite dimensions that summarize the dominant axes of neighborhood differentiation. However, because these components are constrained to be linear combinations of the input variables, there is an inherent assumption that the underlying structure of the data can be adequately described by linear relationships. These linear assumptions can be problematic when applied to spatial data, which frequently exhibit spatial heterogeneity (non-stationary relationships across space) and spatial autocorrelation (nearby areas having similar demographic profiles) (Demšar et al. 2013). Socio-economic and demographic phenomena commonly produce non-linear dependencies, interaction terms, and threshold effects that PCA's orthogonal linear components cannot adequately capture.

Developing summaries of small-area residential structures that better respect the non-linearities and dependence inherent in spatial data remains under-explored. Within the broader family of high-dimensional reduction techniques, manifold-learning methods provide a principled response to this challenge. They rest on the manifold hypothesis, which states that although observations are recorded in a high-dimensional feature space, the salient structure often lies on a much lower-dimensional manifold embedded within it. By learning this geometry, rather than relying on linear projections, such methods aim to uncover meaningful patterns in residential datasets that would otherwise be obscured. A diverse range of non-parametric techniques has emerged, including t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten and Hinton 2008), Uniform Manifold Approximation and Projection (UMAP) (Healy and McInnes 2024), Kernel PCA (Schölkopf et al. 1998), and Diffusion Maps (Coifman and Lafon 2006). Alongside these, parametric neural-network approaches can also learn complex manifold structures. Autoencoders offer powerful unsupervised dimensionality-reduction frameworks (Wang et al. 2016), pairing an encoder that compresses inputs into a low-dimensional latent space with a decoder that reconstructs the original data from that representation (Rezende et al. 2014; Kingma and Welling 2019).

Manifold-learning methods have proved valuable for reducing complex, high-dimensional datasets. In geographic analysis, they are especially prominent in remote sensing (Crawford et al. 2011); for example, variational autoencoders have been used for embedding extraction from satellite imagery (Belwalkar et al. 2018; Singleton et al. 2022). t-SNE and Laplacian Eigenmaps have been used to characterize multi-scale spatio-temporal processes from satellite imagery and raster models (Sousa and Small 2022), and t-SNE has supported cluster analysis and spatial delineation in groundwater geochemistry (Liu et al. 2021). UMAP has also been shown to be a powerful tool for exploring population substructure and fine-scale genetic geographic relationships in human genomic data (Diaz-Papkovich et al. 2019; Diaz-Papkovich et al. 2021). Applications that are related to small-area socio-economic and demographic data remain more limited, though the evidence base is growing. For example, diffusion maps have been used to analyze the 2011 England and Wales Census data (Barter and Gross 2019; Xiu and Chen 2023). Furthermore, domain-specific studies include energy-infrastructure analysis using UMAP and t-SNE on geolocated smart-meter data (Khan et al. 2023) and housing market applications (Chen et al. 2025). While such a body of work demonstrates the potential of manifold learning for geographic analysis, the suitability for operational applications exploring residential differentiation where transparency, updatability, and consistency over time are paramount requires further scrutiny.

1.2 | Practical Requirements for Composite Indicators of Neighborhood Structure

Composite indicators of residential differentiation are often embedded in policy workflows, updated periodically, and subject to stakeholder scrutiny. Such operational demands place specific requirements on dimensionality reduction methods that extend beyond representational fidelity alone. When developing measures of residential differentiation, the application of non-parametric manifold learning techniques does, however, have a number of constraints. Although methods such as t-SNE and UMAP have demonstrated utility as tools for exploratory visualization, they are designed to preserve local neighborhood structure at the expense of global distances, so the reduced embedding does not faithfully retain the relative positions of areas across the dataset as a whole (Sainburg et al. 2021; Healy and McInnes 2024). Beyond this, standard implementations of both methods lack any natural mechanism for inverse reconstruction. This restricts their ability to quantify what information is preserved or lost during reduction, which is an issue of particular concern where socio-economic indicators carry policy weight and transparency may be prioritized. Practical implementation of non-parametric manifold learning methods may also introduce operational hurdles that limit their effectiveness where measures need updating over time. When new input data become available, whether from temporal updates or through the addition of geographic areas, t-SNE requires an entire dataset to be reprocessed. For example, if a single new area was included in an analysis, this could necessitate re-analyzing thousands of existing areas, with no guarantee that the resulting patterns will

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

align with previous findings. UMAP provides a partial solution through an approximate projection method for new data points, yet such an approach sacrifices accuracy and tends to not fully replicate the quality achieved through complete retraining of the model. Furthermore, neither Kernel PCA nor Diffusion Maps naturally accommodate new data points without substantial recomputation. While approximate out-of-sample extensions exist (Williams and Seeger 2000; Coifman et al. 2005), such methods can introduce approximation errors and also require retention of the original training data for projection computations. Finally, sensitivity to parameter selection presents another practical obstacle for many methods. Both t-SNE and UMAP require researchers to specify parameters that fundamentally shape the resulting analysis. For example, perplexity values in t-SNE determine the balance between local and global structure preservation, while UMAP's neighborhood and distance parameters control how tightly clusters form in the reduced space. Such parameters may require different values when analyzing urban versus rural areas, or when examining data at different geographic scales, making it difficult to establish standardized global analytical approaches. Kernel PCA requires selection of both kernel type (e.g., Gaussian, polynomial, and sigmoid) and associated parameters, with the Gaussian kernel's bandwidth parameter proving particularly influential in determining the scale at which demographic patterns are detected. Diffusion Maps also face similar challenges with their scale parameter, which controls the diffusion process that fundamentally shapes the resulting geometric representation. Although parametric variants of t-SNE and UMAP have been proposed to enable mappings for new data and partial interpretability (van der Maaten 2009; Sainburg et al. 2021), these techniques remain less widely adopted and introduce additional model complexity.

We argue that autoencoder (AE) architectures offer a practical and maintainable alternative for characterizing residential structure that addresses several constraints of non-parametric manifold methods. First, AEs provide a parametric encoder, enabling out-of-sample projection of new small areas without necessity to recompute existing embeddings; and as such, models can be fine-tuned when the geography or temporal update allows, while keeping earlier embeddings valid. Second, the decoder enables reconstruction diagnostics, allowing us to quantify what information is preserved or lost and to relate latent dimensions back to the original variables via variable-wise reconstruction error and sparse surrogate models. We see this as an advantage for policy-facing applications and transparency. Third, representation learning reduces reliance on fragile hand-tuned manifold parameters; instead, the network learns non-linear structure directly from the data. Prior work already points to their utility for geodemographic preprocessing (De Sabbata and Liu 2023; Liu and De Sabbata 2019), the construction of geographic embeddings for zonal units (Cepeda et al. 2023; Agarwal et al. 2024), and the derivation of built-environment structure from satellite imagery (Singleton et al. 2022; Klemmer et al. 2024).

Building on this evidence, we develop an AE-based pipeline for deriving a composite indicator of residential differentiation using 2021 Census small-area data for England and Wales. We then evaluate this against PCA as a linear benchmark. Within this context, our paper contributes to the literature on residential differentiation by demonstrating how autoencoders can offer a robust and effective alternative to traditional linear methodologies in deriving composite indicators. We showcase the efficacy of this new approach through a detailed, extensive case study evaluation utilizing neighborhood-level data from the 2021 Census for England and Wales. We argue that AEs, when applied for a similar purpose to PCA as a dimensionality reduction method, provide additional advantages by effectively capturing localized non-linear relationships inherent in geographic data, directly addressing a key limitation of PCA.

2 | Autoencoder-Based Census Dimensionality Reduction

Inputs to this study were all extracted from the 2021 England and Wales Census and downloaded in bulk from Nomis (https://www.nomisweb.co.uk/sources/census_2021_bulk) at the Output Area (OA) scale. An OA is the smallest unit for which census data in the UK are disseminated and typically comprises between 40 and 250 households. Data for England and Wales were used in this study, which represent 188,880 OAs. Topics that were extracted included Demography and Migration, Ethnicity, Identity, Language and Religion, Work and Travel, Housing, Education, and Health. Variables with limited geographic coverage were excluded, including Welsh language (available only for Wales), short-term non-UK residents, and second home type (both containing non-zero values for only a minority of OA). Population density was also excluded. This resulted in a subset of 408 initial variables, all presented as percentages relative to their appropriate table denominator. The initial inputs therefore represented a very large multidimensional array (77,063,040 data points) and contained measures that covered the full breadth of topics within the Census at OA scale, with full coverage over both England and Wales.

2.1 | Autoencoders

The structure of an autoencoder comprises two primary stages: the encoder and the decoder (Figure 1). The encoder condenses input data into a more compact representation, while the decoder aims to reconstruct the original data from the condensed form. The autoencoder's architecture features a sequence of densely connected layers that narrow, culminating in a "bottleneck" layer which is the most compressed state. Autoencoders learn hierarchical features from the data, which can be far more expressive than the linear combinations given by PCA. A deep autoencoder can stack multiple layers that capture first simple patterns, then more abstract ones. For instance, in image processing, early layers typically detect low-level features such as edges and textures, while deeper layers capture higher-order representations like object parts or spatial arrangements. Analogously, when applied to spatial tabular data such as census variables, autoencoders can learn composite latent factors that encode co-occurring characteristics, for example, combining patterns of population aging with high ethnic diversity into a single learned representation.

There are multiple ways in which an autoencoder can be designed, with options including the number of layers, the types of activation functions used and the choice of loss function. The selection of these elements can influence the performance and

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

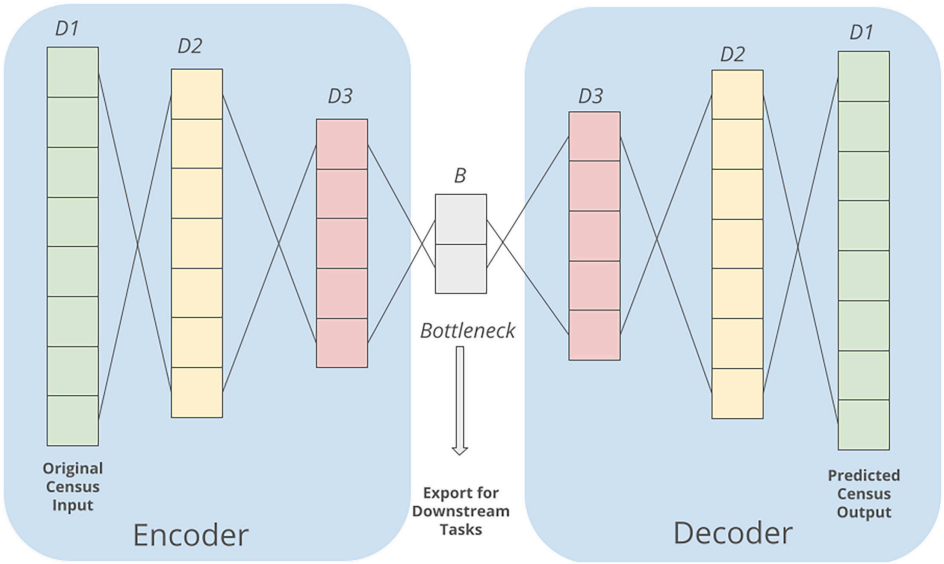


FIGURE 1 | The model encoder-decoder architecture. D1, D2, D3, and B indicate the layer dimensions, with D1 the input and output layer, D2 and D3 the hidden layers, and B the bottleneck that sets the latent dimensionality.

capability of the autoencoder. The number of layers determines the depth and complexity of the network architecture. Deeper networks with more layers can learn increasingly abstract representations of the input data, while shallower architectures may be sufficient for simpler patterns. The bottleneck layer, positioned at the network's center, defines the dimensionality of the compressed representation and thus the degree of information reduction. Activation functions can introduce non-linearity into the network, enabling it to capture complex, non-linear relationships in the data. Common choices include ReLU (Rectified Linear Unit), sigmoid, or tanh functions. Without such non-linear activation functions, the autoencoder would be limited to a linear transformation that recovers the same subspace as PCA (Baldi and Hornik 1989). The model is operationalized against a loss function which evaluates how the model's accuracy changes during training, essentially measuring error related to the reconstruction of the original source data. By training to minimize this reconstruction error, the autoencoder focuses on preserving essential structure, potentially yielding features that align better with meaningful geographic patterns, at the cost of being less directly interpretable than PCA components.

2.2 | Model Architecture and Training

For this implementation we designed an AE with a depth of three dimensionality-reducing layers, each followed by a Leaky ReLU activation function that aimed to introduce non-linearity into the model. The model architecture is shown in Figure 1 where D1, D2, D3 signify different dimensions resulting in the most compressed state at B, the "bottleneck" within the autoencoder. D1 is fixed by the data as the 408 input variables and is also the dimensionality of the reconstructed output. D2 and D3 are the hidden layers, with dimensions decreasing linearly in three steps from D1 to B. Single lines between each layer are illustrated, but in reality the connections are dense, as their number is the product of the number of neurons in the preceding layer and the number of neurons in the subsequent layer.

The model was implemented with a loss function comprising the mean squared error (MSE), normalized by the mean squared value of the input. Squaring the error penalizes larger deviations more heavily. The normalization ensures the loss is scale-invariant, preventing variables with larger magnitudes from dominating the training process. The autoencoder is trained for a series of epochs, where an epoch represents each training cycle. The MSE within each epoch can be monitored until further iterations result in very limited improvement to the loss function. Although it was possible to utilize the whole dataset for training, a "mini-batch" approach was implemented instead. This involved splitting the dataset into small (1% in our case) and randomly chosen subsets. All subsets were then processed separately during each epoch, thus updating the model parameters multiple times (100 times in our case) per epoch, rather than just once using the entire dataset. This also resulted in the returned models showing marginally better MSE scores. This occurs because processing smaller subsets of data multiple times per epoch provides more frequent and stable gradient updates which helps the model to avoid local minima. To assess training stability, we trained each model configuration 10 times with different random initializations. The models were run on a pair of NVIDIA RTX A6000 Nvidia GPUs, using Pytorch and Pytorch Lightning with CUDA enabled. Each model trained in under 10 min with a storage footprint under 2 MB. Table 1 contains the configuration and hyperparameters of the AE models used.

3 | Evaluating Non-Linear Information Gain

A key contribution of this paper is to examine the extent to which non-linear geographic information can be captured by the composite measures derived from AEs. As such, we evaluated the performance of the AE in creating a compressed representation of the England and Wales Census data by comparing it to a PCA decomposition across a series of latent dimensionalities, ranging from 2 to 128. The model architecture, as described above, is kept

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Autoencoder parameter settings.

Parameter	Parameter setting
Bottleneck dimensionality (B)	2, 4, 8, 16, 32, 64, 100, 128
Architecture	Symmetric encoder-decoder (D1, D2, D3, B)
Layer sizes	Linear from 408 to B (e.g., D1 = 408, D2 = 274, D3 = 141, B = 8)
Hidden activation	LeakyReLU ($\alpha = 0.01$)
Optimizer	AdamW (lr = 10^{-3}, weight decay = 10^{-2})
LR schedule	ReduceLROnPlateau (factor = 0.2, patience = 10)
Loss function	Normalized MSE
Epochs	500
Batch size	1% of observations
Random seed	$20210321 + (i \times 1000)$ for $i \in \{0, 1, \ldots, 9\}$

the same for each training run, with the layer size configured to decrease linearly from the input layer to the bottleneck size.

The linear comparator was generated by calculating the reconstruction error of the PCA decomposition using the number of principal components (ordered by explained variance) corresponding to each AE latent dimensionality. PCA was applied to the same 408-variable input used to train the autoencoder, expressed as percentages relative to their appropriate table denominator. Reconstruction is achieved by projecting the data into the reduced space and applying the inverse transformation back to the original feature space. This reconstruction represents the optimal linear approximation of the data for a given number of components. Any residual error, therefore, reflects variance that cannot be captured by that number of linear combinations of the original variables.

As discussed in Section 1.2, non-parametric methods such as UMAP provide no natural inverse transformation to the original feature space, so they are not included in the comparison. The commonly used approximate inverse reconstructs a point by interpolating between the original training observations at the vertices of a Delaunay triangulation of the embedding, so it is not self-contained. Furthermore, the cost of computing this inverse grows exponentially with the latent dimension, so it becomes intractable for our purposes, failing beyond eight dimensions. The parametric variant of UMAP learns the embedding with a neural network and can be paired with a neural decoder to reconstruct the inputs, but this makes it an autoencoder-hybrid model that overlaps with the approach evaluated here.

We initially assessed the performance of the respective methods using the root mean squared error (RMSE) between the reconstructed and the original census data, computed across all variables and Output Areas (OAs). Taking the square root of the mean squared error returns the metric to the original units of the input data, which in this case are percentages. Variability in reconstruction error across runs was minimal, with the coefficient of variation (the ratio of standard deviation to mean RMSE) below 0.014 for all tested dimensionalities. This demonstrates stable training behavior and consistent performance, confirming that differences between AE and PCA are robust.

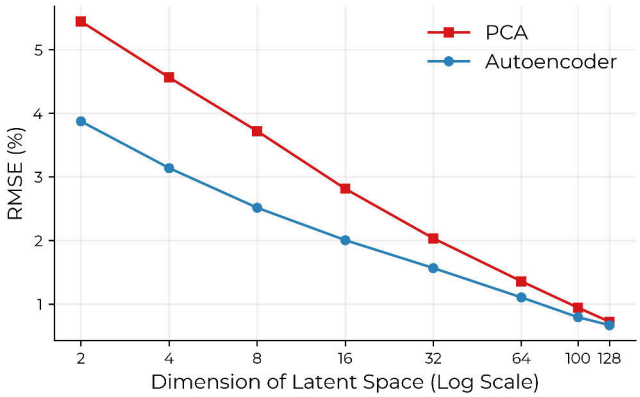


FIGURE 2 | A comparison of the reconstruction error between PCA and autoencoder methods as a function of compressed latent space dimensionality. Autoencoder results represent the mean RMSE over 10 independent training runs, with between-run variability too small to be visible at this scale.

Figure 2 illustrates the outcome of this comparison. For both PCA and AE, as the latent dimensionality increases, the RMSE decreases, as the compressed representation contains more information and therefore produces a better reconstruction of the original data. The AE outperforms PCA in reconstruction accuracy for all tested dimensions, demonstrating its ability to capture non-linear structure in the data that the linear decomposition cannot. The performance gain is most pronounced at lower dimensionalities. As the latent dimensionality approaches that of the original input data, the advantage of non-linear methods diminishes: with minimal compression, there is less complex structure for the autoencoder to exploit, and both methods converge toward near-perfect reconstruction.

3.1 | Out-of-Sample Generalization

The reconstruction errors reported above are measured on the same OAs used to fit each model. As the autoencoder has significantly more free parameters than PCA's linear projection, the lower reconstruction error could in principle reflect its greater capacity to fit the training areas rather than a genuinely more efficient encoding of the data. Although the low-dimensional bottleneck constrains capacity, a sufficiently flexible encoder and decoder can still fit structure specific to the training areas rather than structure shared across them. We use a spatially independent hold-out set to assess how well each model encodes and reconstructs areas that were not used in training.

Census variables are spatially autocorrelated, so a random OA-level hold-out risks leakage between neighboring training and test areas. We therefore withhold whole Local Authority Districts (LADs), which nest OAs into 331 larger contiguous areas.

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Reconstruction RMSE (%) under 5-fold spatially blocked cross-validation.

Dimensionality	PCA (train)	PCA (held-out)	AE (train)	AE (held-out)
2	5.445	5.465 (5.335–5.539)	3.909	4.071 (3.999–4.114)
4	4.565	4.580 (4.498–4.677)	3.134	3.328 (3.273–3.385)
8	3.720	3.744 (3.683–3.846)	2.511	2.648 (2.593–2.681)
16	2.813	2.836 (2.744–2.897)	2.003	2.084 (2.024–2.112)
32	2.033	2.055 (1.970–2.110)	1.562	1.607 (1.560–1.634)
64	1.359	1.380 (1.317–1.413)	1.105	1.134 (1.098–1.165)
100	0.944	0.964 (0.915–0.989)	0.787	0.813 (0.770–0.870)
128	0.723	0.745 (0.704–0.780)	0.685	0.708 (0.665–0.792)

Note: Held-out values are the mean (min–max) across the five held-out folds; train values are the mean across the five training fits.

The 331 LADs are partitioned into five folds of entire districts, each containing ∼20% of OAs (37,313–38,014 of 188,880), and each fold is withheld in turn. Using the same architecture and hyperparameters as in Table 1, the autoencoder is trained five times, once per fold, each time on the remaining 80% of areas and evaluated on the OAs in the held-out districts. The PCA comparison is likewise refit and evaluated on the same five train and test partitions. Table 2 reports reconstruction RMSE on the training partition ("train") and on the withheld districts ("held-out"), with the range (min–max) across the five held-out folds.

The training-sample RMSE values from the five spatial hold-out fits agree closely with the full-sample RMSE values reported above, differing by less than 0.04 percentage points at every dimensionality. This indicates that the refitted models are consistent with the models trained on the complete dataset.

The autoencoder retains its advantage over PCA on the out-of-sample areas at every dimensionality. Both models exhibit some deterioration in error on the held-out areas. For PCA, the train–held-out gap is almost constant at approximately 0.02 percentage points across dimensionalities, providing a useful baseline for the change in error introduced by evaluating on spatially separate areas. The autoencoder has a larger generalization gap, particularly at lower dimensionalities. Across the more compressed representations, held-out RMSE exceeds train RMSE by up to 0.19 percentage points, equivalent to a relative increase of approximately 6%. This gap falls as dimensionality increases. At dimensions 32 and above, those most likely to be used in practice, the autoencoder's held-out RMSE exceeds its training RMSE by less than 0.05 percentage points, corresponding to a relative increase of less than 3.5%. In these cases, the out-of-sample fold ranges overlap the mean training RMSE, placing the train–test difference within the scale of fold-to-fold variation.

These results address the concern that the AE's advantage over PCA could simply reflect its greater capacity to fit the training OAs. Were that the case, we would expect its error to increase substantially on held-out areas. Instead, the gap between training and held-out error remains small, particularly at the higher dimensionalities most relevant for subsequent analysis. The remaining evaluations therefore use models trained on the complete dataset, so that the substantive analyses are based on representations derived from the full set of OAs.

TABLE 3 | Global Moran's I by latent dimensionality.

Dimensionality	PCA	Autoencoder (mean ± SD)
2	0.536	0.407 ± 0.008
4	0.515	0.362 ± 0.008
8	0.528	0.316 ± 0.006
16	0.467	0.310 ± 0.003
32	0.444	0.355 ± 0.003
64	0.458	0.419 ± 0.002
100	0.494	0.474 ± 0.003
128	0.495	0.473 ± 0.008

Note: All p-values < 0.001.

4 | The Geography of Reconstruction Errors

This section examines the spatial structure of the reconstruction errors. We consider the RMSE per Output Area to better understand within which geographic contexts each model struggles to reconstruct the original data, and whether such residual structure aligns with known socio-spatial patterns. Initially we examine spatial autocorrelation and then explore these patterns across both deprivation and density gradients, and across geodemographic clusters. Finally, we map the reconstruction errors to visualize these spatial patterns directly.

A Global Moran's I for each latent dimensionality is presented in Table 3, computed using a queen contiguity spatial weights matrix of Output Areas. All values were positive and statistically significant ($p < 0.001$), indicating spatial clustering of reconstruction errors for both methods. However, the autoencoder consistently produced lower Moran's I values across all dimensionalities, suggesting it captures more of the spatially structured variance in the original data, leaving residuals that are more spatially random. This indicates that the autoencoder learns representations that better preserve the geographic coherence of neighborhoods compared to PCA. Neither method exhibited a clear monotonic relationship between dimensionality and spatial autocorrelation, though the autoencoder showed a slight U-shaped pattern with minimum values at intermediate dimensions (8–16). The persistent spatial structure in reconstruction errors for both approaches

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

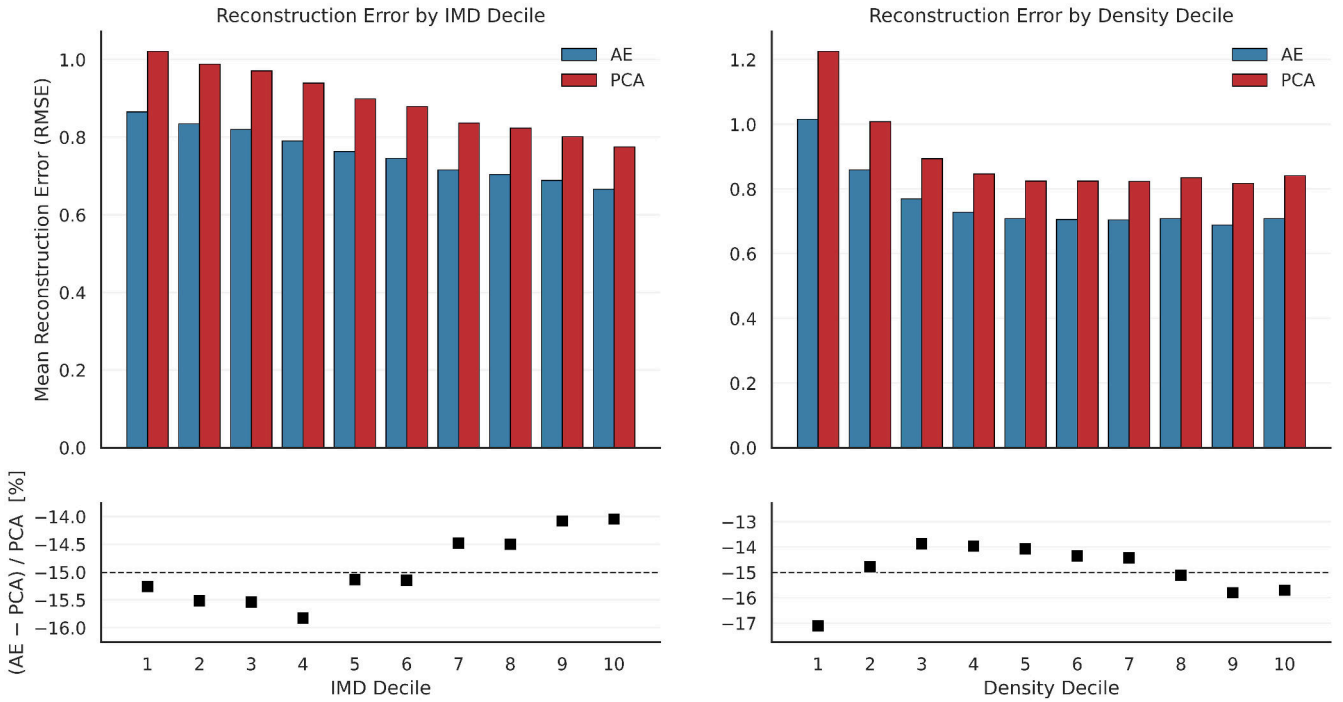


FIGURE 3 | Reconstruction error distribution by IMD (1 most deprived, 10 least deprived) and population density decile (1 most dense, 10 least dense). The dashed line indicates the mean improvement across deciles.

suggests that some geographic patterns remain uncaptured in the latent representations, potentially reflecting unmeasured local processes or spatial dependencies not fully encoded into the input features.

To further evaluate and compare reconstruction scores, we selected a latent dimensionality of 100 for the AE and 100 components for PCA, a typical dimensionality for applications using reduced census data as input (Liu et al. 2019). Given the low variance in AE reconstruction performance across training runs, we use the median-performing model for all subsequent spatial and geodemographic analyses. For this dimension size, the mean reconstruction RMSE across Output Areas is 0.943 for PCA and 0.760 for the median AE run, a relative improvement of 19.4%. We then categorized or derived measures for each OA using several metrics: population density, Index of Multiple Deprivation (IMD),[1] and the 2021 Output Area Classification (2021 OAC; Wyszomierski et al. 2023) in order to systematically analyze the characteristics of the error distribution. This identified which types of neighborhood were most accurately represented and which suffered the greatest information loss in each compression method.

The first two of these measures are external to the model inputs. Population density is independent by construction, since the model was trained only on proportional census variables and absolute density was not an input. The IMD is the official measure of relative deprivation for small areas, built from weighted domains and compiled predominantly from administrative data, with only a small number of census-derived indicators. The distribution gradients, shown in Figure 3, are broadly similar for both PCA and the AE, with higher errors associated with more deprived and more densely populated areas. Reconstruction error is 31.8% higher for the most deprived decile compared to the least deprived for PCA, and 29.9% higher for AE. A similar pattern is observed for population density, where error is 45.7% and 43.3% higher in the most densely populated decile compared to the least dense for PCA and AE, respectively. Across all indices, the AE consistently yields lower reconstruction errors than PCA. However, the most substantial relative improvements are observed in OA nested within Lower layer Super Output Area (LSOA) with lower IMD deciles 1–4, highlighting its increased effectiveness in capturing patterns in more materially deprived areas. Likewise, the greatest reductions in error for population density occur in the highest density decile. This suggests that the autoencoder not only outperforms PCA in terms of overall reconstruction accuracy but also helps mitigate an unequal distribution of error across socio-economic and spatial gradients.

The 2021 OAC is a hierarchical geodemographic classification developed in partnership with the Office for National Statistics that identifies OAs with similar population and built environment characteristics (Wyszomierski et al. 2023). Constructed using a subset of 2021 Census variables, the classification assigns OAs to one of 8 Supergroups, which subdivide into 21 Groups and 52 Subgroups. Table 4 contains a list of the Supergroups and Groups produced by the classification; Subgroups are not used in this study. Full cluster descriptions and statistics can be accessed elsewhere.[2]

As the OAC is constructed using a subset of 2021 Census variables, some overlap exists with the variables used to train the autoencoder. We therefore use it here as a descriptive way to compare reconstruction error across recognizable area profiles, rather than as an independent validation. As an established classification of area types, the OAC provides a useful basis for assessing

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 | 2021 OAC Supergroup and Group names.

Supergroup	Group
1: Retired Professionals	1a: Spacious Rural Living
	1b: Small Town Suburbia
	1c: Established Mature Families
2: Suburbanites and Peri-Urbanites	2a: Inner Suburbs and Small Town Living
	2b: Rural Amenity
	2c: Aging Communities
3: Multicultural and Educated Urbanites	3a: Student Living and Professional Footholds
	3b: Ethnically Diverse Young Families
	3c: Diverse Educated Urban Singles
4: Low-Skilled Migrant and Student Communities	4a: Ethnically Diverse Families in Less Connected Locations
	4b: Established Multi-Ethnic Communities
	4c: Challenged Multicultural Communities and Students
5: Ethnically Diverse Suburban Professionals	5a: Outer Suburbs
	5b: Suburban Professionals
6: Baseline UK	6a: Challenged Communities
	6b: Legacy Industrial and Coastal Communities
	6c: Multicultural Inner Suburbs
7: Semi and Un-Skilled Workforce	7a: Established but Challenged
	7b: Young Families in Industrial Towns
8: Legacy Communities	8a: Routine Occupations or Retirement
	8b: Legacy and Demographically Mixed Communities

whether reconstruction error differs across different profiles of places present in the census data.

Figure 4 displays the reconstruction error for the PCA and AE, segmented by OAC Supergroup and the more granular Groups. The distribution of reconstruction errors for both PCA and the autoencoder is not uniform across Supergroups. Supergroups 3 and 4 exhibit the highest overall errors, indicating greater complexity or heterogeneity in these population clusters. Notably, within these supergroups, Groups 3a (Student Living and Professional Footholds) and 4c (Challenged Multicultural Communities and Students) show the greatest improvement in reconstruction error when using the autoencoder compared to PCA. This suggests that the autoencoder better captures the complex and non-linear structure of these areas. These Groups are characterized by highly variable demographic, housing, and socio-economic patterns associated with transient populations, student accommodation, and diverse communities; patterns that are less well captured by the linear assumptions of PCA but more effectively modeled by the non-linearity and capability of the autoencoder.

To further explore the geography of error, Figure 5 maps the spatial distribution of reconstruction errors for both PCA and AE, aggregated up to the level of Middle Layer Super Output Areas (MSOAs) for visual clarity. Each MSOA comprises approximately 26 Output Areas (OAs). The density-related trends described earlier are clearly reflected, with higher reconstruction errors concentrated in urban areas. This is most apparent in London, the most densely populated city in the UK. Additional geographic outliers in rural regions such as Cornwall and North Wales can be observed distinctly in the PCA map. Here poorer reconstruction of variables regarding national identity contributes significantly to this error. The Cornish and Welsh identities are largely unique to these parts of the country and therefore have non-linear relationships with other census variables, which PCA fails to capture effectively. The AE has a lower RMSE than PCA for all 7264 MSOAs, indicating that AE outperforms PCA almost universally.

5 | Case Study: Using the AE Embeddings to Describe Salient Neighborhood Clusters

The preceding sections have demonstrated that the autoencoder produces latent representations with lower reconstruction error and greater spatial coherence than PCA across all tested dimensionalities. These findings establish that the learned embeddings encode meaningful geographic structure; however, the practical value of such representations ultimately depends on their utility within downstream analytical tasks. In this section, we evaluate one such application by examining whether the AE latent space can serve as a robust basis for geodemographic classification, a common and policy-relevant use case for dimensionality-reduced census data.

5.1 | Clustering Methodology

To assess the capacity of the AE embeddings to recover coherent neighborhood typologies, we applied k-means cluster

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

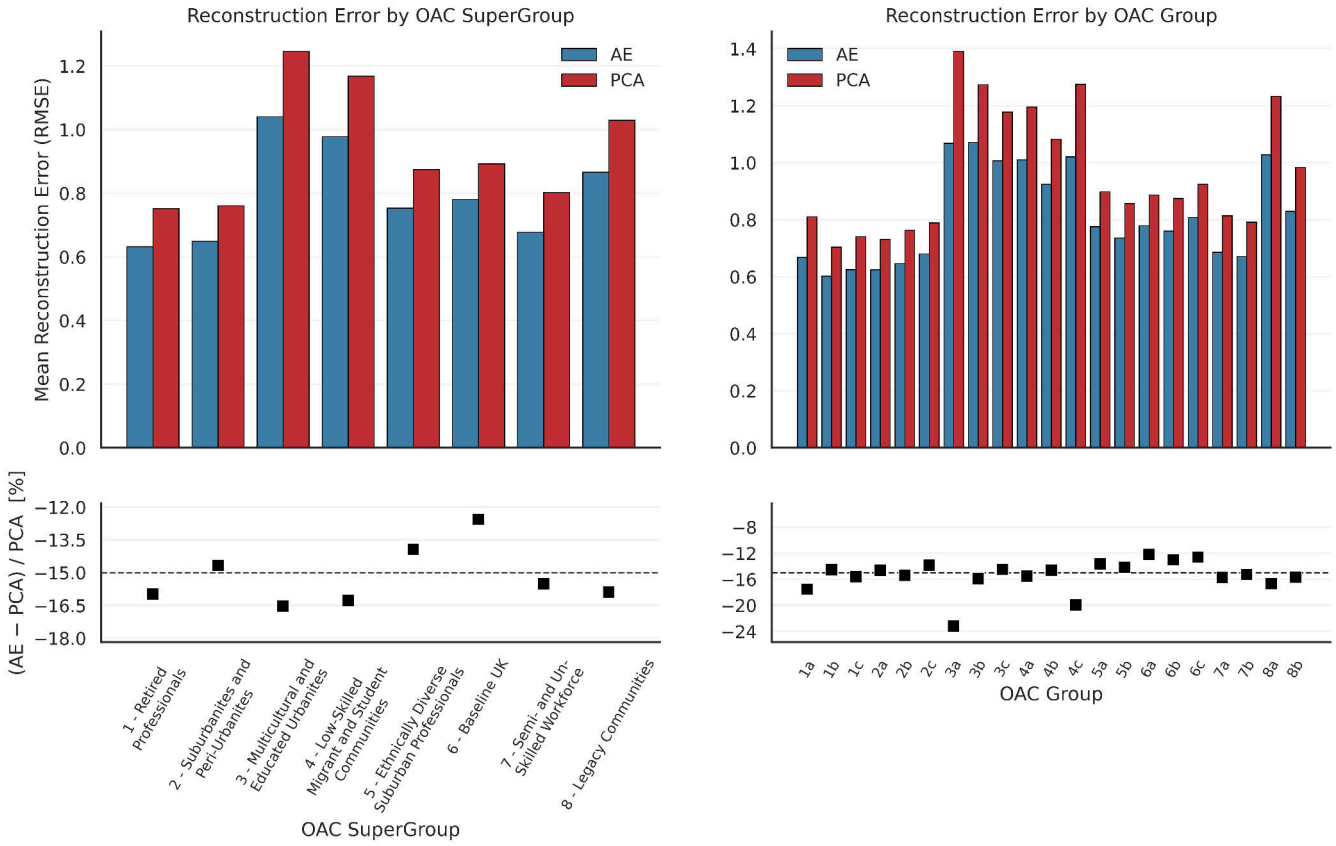


FIGURE 4 | Reconstruction error distribution by OAC Supergroup and OAC Group. The dashed line indicates the mean improvement across Supergroups and Groups.

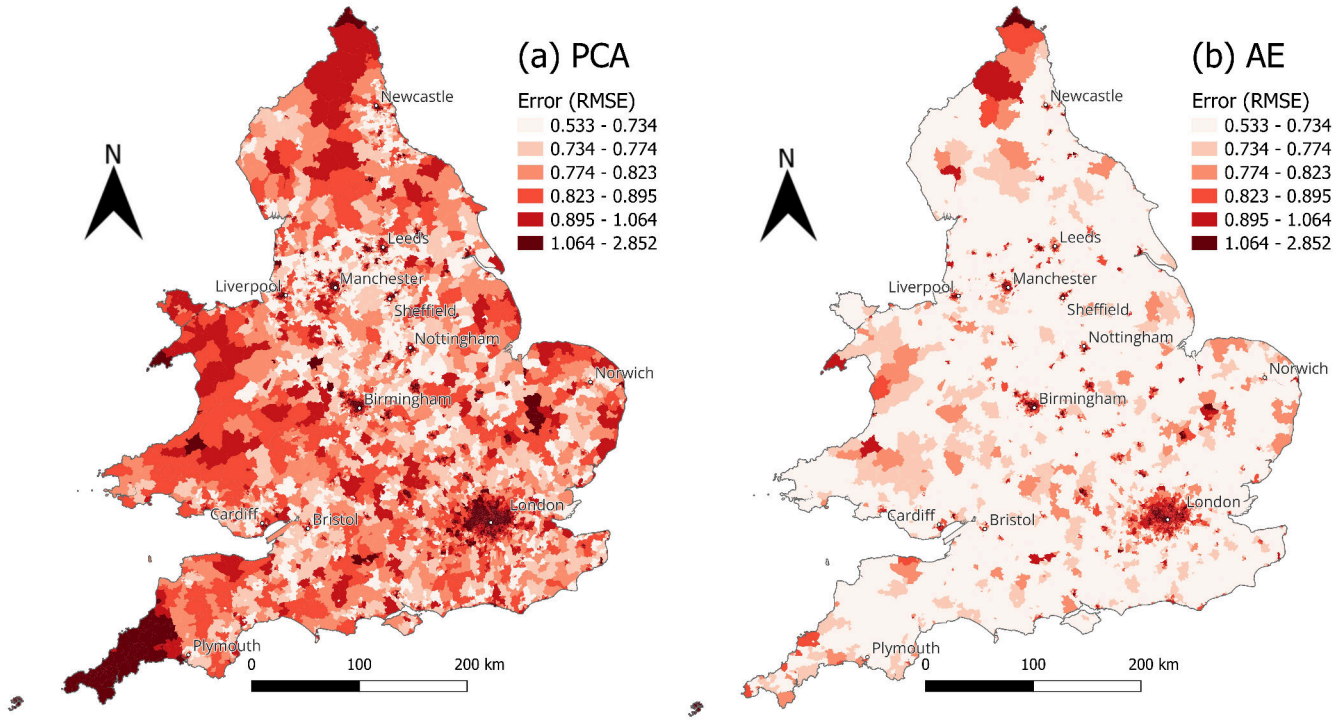


FIGURE 5 | The mean reconstruction error for each MSOA in England and Wales for (a) the PCA reconstruction with 100 components, and (b) the Autoencoder reconstruction with a latent dimensionality of 100.

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

analysis to the 100-dimensional latent space. This dimensionality was selected to maintain consistency with the evaluation framework established in Section 4, where 100 components represent a typical operational specification for census-based applications (Liu et al. 2019). The number of clusters was set to $k = 8$, determined through inspection of a Clustergram, following the methodological conventions employed in the construction of the 2021 Output Area Classification (Wyszomierski et al. 2023). To ensure stability of the final solution, the algorithm was executed with 10,000 random initializations, with the configuration yielding the lowest within-cluster sum of squares retained. An equivalent clustering procedure was applied to the 100-component PCA representation to enable direct comparison of the two dimensionality reduction approaches as feature extraction stages for neighborhood classification. In order to compare the results of the two clustering outcomes, the cluster labels were aligned by pairwise matching between AE and PCA clusters with the largest OA overlap. Under this alignment, 84.3% of OAs were classified identically by the AE- and PCA-based solutions. The Adjusted Rand Index quantifying the overlap between the two solutions was 0.68, indicating both solutions produce similar geodemographic structure.

5.2 | Spatial Coherence of AE-Derived Clusters

The resulting classifications produce spatially coherent neighborhood groupings (Figure 6), with clear geographic structure visible at both national and metropolitan scales. Both the AE latent space and the PCA components produce clusters with strong spatial contiguity despite the absence of any explicit spatial constraint in the clustering algorithm, indicating each encodes geographic variation in a manner that naturally groups proximate areas with similar socio-demographic profiles. Non-linear compression therefore does not come at the cost of spatial regularity, which is a desirable property for policy-oriented geodemographic systems, where classifications are expected to reflect recognizable local contexts rather than spatially fragmented artifacts of the dimensionality reduction process.

To evaluate whether the AE-derived clusters capture substantively meaningful population structure, we examined the correspondence of both solutions with the 2021 OAC Supergroups. The 2021 OAC represents an established geodemographic framework constructed through conventional methods, including variable selection, standardization, and iterative k-means clustering. As it is derived from the same 2021 Census inputs used here, it provides a validated benchmark method against which the groupings can be compared. Figure 7 presents a cross-tabulation of OA frequencies between OAC Supergroups and the eight clusters for AE and PCA.

Given the substantial overlap between the two clustering solutions, their correspondence with the OAC is broadly similar. Several notable patterns emerge. The Legacy Communities Supergroup maps almost entirely to a single cluster (Cluster 8) under both the AE (92.7%) and PCA (94.7%), indicating that both representations identify a distinct population archetype corresponding to areas characterized by aging populations, routine occupations, and demographic stability. The Semi- and Un-Skilled Workforce Supergroup exhibits a near-identical dominant mapping to Cluster 2 (73.5% AE, 74.4% PCA). Agreement is not confined to Supergroups with a single dominant cluster. The Ethnically Diverse Suburban Professionals Supergroup, for example, divides principally across Clusters 1, 3, and 5 under both solutions, in closely comparable proportions.

Disagreement between the two solutions is concentrated in the Retired Professionals and Suburbanites and Peri-Urbanites Supergroups, of which 23.1% and 23.3% of OAs, respectively, change label, above the national rate of 15.7%. Retired Professionals is principally divided across three clusters under both solutions (Clusters 1, 5, and 6), but the balance differs, with the AE placing 37.5% in Cluster 6 against 18.1% under PCA and the remainder shifting into Clusters 1 and 5. Suburbanites and Peri-Urbanites shows the same loss from Cluster 6, falling from 26.3% under the AE to 15.2% under PCA, though here the mass is absorbed principally into Cluster 3. This is where divergence would be expected. Retired Professionals are distributed throughout England and Wales, and Suburbanites and Peri-Urbanites across suburbs and rural fringes nationwide, rather than being concentrated in particular settings. Neither Supergroup is sharply bounded in the census attributes on which the classification is built, and boundaries drawn through gradual variation of this kind therefore depend on how those attributes are compressed, which linear and non-linear methods do differently.

5.3 | Cluster Quality Benchmarking

To assess cluster quality, we compared the AE-based and PCA-based classifications using two standard internal validation metrics that capture complementary properties of cluster geometry. The AE achieved a higher Calinski-Harabasz index (AE: 27,293; PCA: 26,125), reflecting a slightly stronger global ratio of between-cluster to within-cluster dispersion, whereas PCA achieved a higher silhouette score (AE: 0.111; PCA: 0.125), reflecting a marginally clearer separation between each area's assigned cluster and its nearest alternative. The two metrics therefore favor different methods, and both differences are small. We interpret modest differences in metrics as indicating broadly comparable classification quality across the two approaches.

5.4 | Decoder Validation of Cluster Centroids

A distinctive advantage of the autoencoder framework is that the trained decoder can be used to project cluster centroids from latent space back into the original variable space, yielding interpretable neighborhood profiles for each cluster. To assess the fidelity of this inverse mapping, we computed the centroid of each cluster in the 100-dimensional latent space, decoded it through the trained decoder network, and compared the resulting reconstructed variable profile against the actual mean of census variables for all Output Areas assigned to that cluster. The decoded centroids matched actual cluster means with an RMSE below 0.3 percentage points. As the decoder is non-linear, decoding this averaged point is not equivalent to averaging the members' profiles in the original variable space, so the close agreement is not guaranteed by the training objective. It instead indicates that each cluster occupies a region of the latent space coherent

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

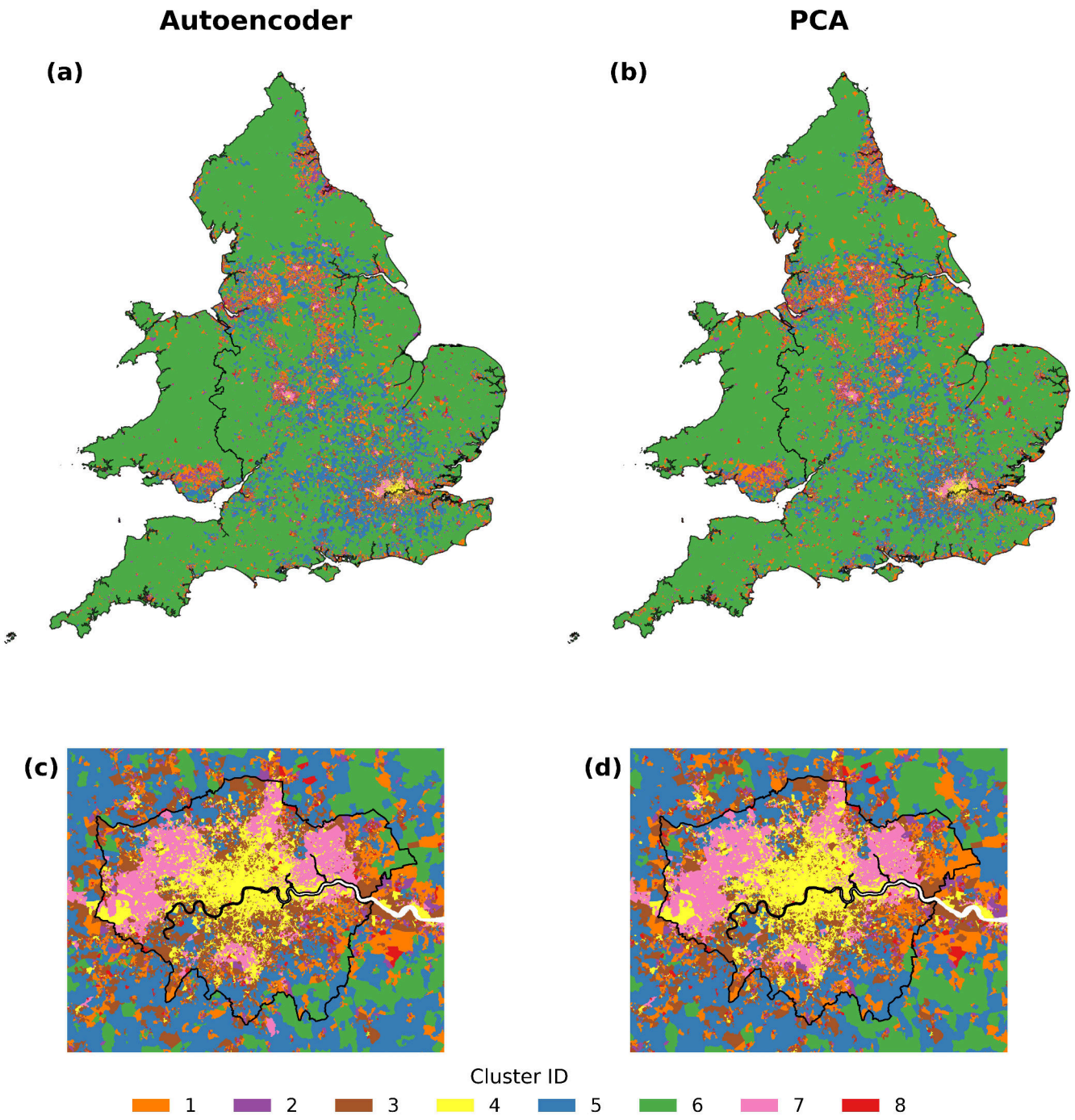


FIGURE 6 | Geographic distribution of the eight cluster geodemographic classification produced by k-means clustering ($k = 8$) on the 100-dimensional compressed space, shown for the AE-derived classification in England and Wales (a) and Greater London (c), with the equivalent PCA-derived classification in (b) and (d).

enough for its centroid to decode to a neighborhood profile representative of its members. This result has important practical implications: it demonstrates that the decoder provides a reliable mechanism for translating abstract latent representations into policy-legible descriptions expressed in the original census variable space. Stakeholders can therefore examine what each cluster "means" in terms of concrete demographic, socio-economic, and housing characteristics without requiring direct access to the latent space itself. This addresses one of the principal concerns raised regarding the application of neural network approaches in operational geodemographic contexts, where transparency and communicability of results are essential requirements.

6 | Discussion and Conclusion

6.1 | Summary of Findings

This study set out to evaluate autoencoders as a non-linear alternative to PCA for deriving composite measures of neighborhood

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

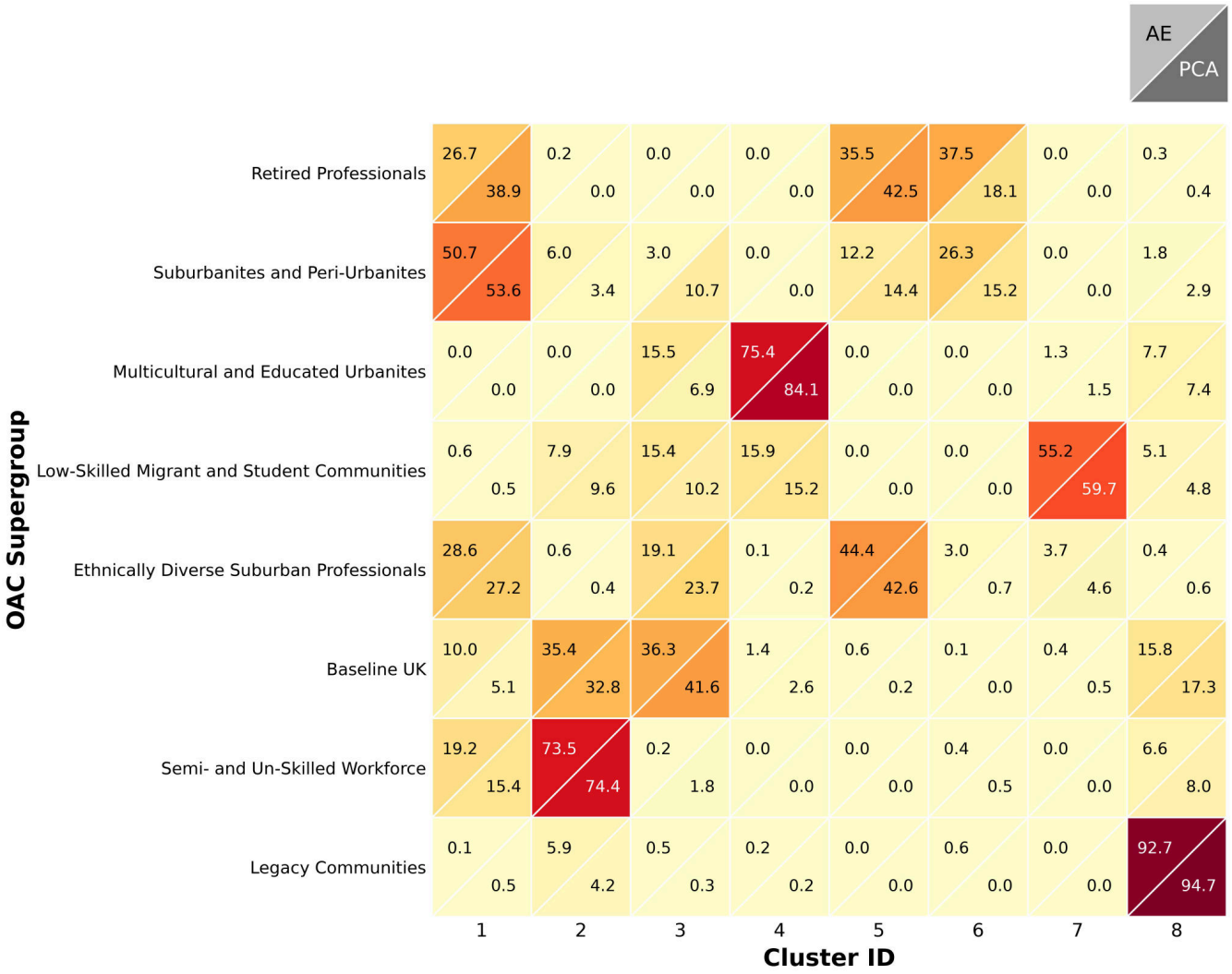


FIGURE 7 | Cross-tabulation of the OA frequency (%) between OAC Supergroups and the 8 clusters produced for both AE (top left) and PCA (bottom right).

structure from census data. The results provide consistent evidence that autoencoders capture information that linear methods do not. Across all tested latent dimensionalities, from 2 to 128, the autoencoder achieved lower reconstruction error than PCA, with the performance differential most pronounced at lower dimensionalities where compression is greatest. At 100 dimensions, a specification typical of operational census-based applications, the autoencoder reduced mean reconstruction RMSE by 19.4% relative to PCA. Training was stable across 10 independent initializations, with a coefficient of variation below 0.014 at all dimensionalities, indicating that the observed improvements are robust to random seed selection rather than artifacts of favorable initialization.

Spatial analysis of reconstruction errors reinforced these aggregate findings. The autoencoder produced residuals with consistently lower spatial autocorrelation (Global Moran's I) than PCA at every dimensionality tested, indicating that it captures a greater share of the spatially structured variance in the original data. When reconstruction errors were stratified by deprivation, population density, and geodemographic classification, the autoencoder outperformed PCA across all categories. The largest relative improvements were concentrated in the most deprived areas (IMD deciles 1–4), the most densely populated areas, and OAC Supergroups characterized by complex demographic composition, notably Supergroup 3 (Multicultural and Educated Urbanites) and Supergroup 4 (Low-Skilled Migrant and Student Communities). These patterns are consistent with the theoretical expectation that non-linear relationships are most prevalent in neighborhoods where multiple demographic processes interact simultaneously.

The case study application in Section 5 demonstrated that the autoencoder's latent representations also serve as effective inputs for downstream geodemographic classification. K-means clustering of the 100-dimensional latent space recovered spatially coherent neighborhood groupings with substantial correspondence to the 2021 OAC Supergroups. Measures of cluster quality were broadly comparable between the AE- and PCA-based classifications, with the AE achieving a higher Calinski-Harabasz index and PCA achieving a higher silhouette score. Decoded cluster centroids matched actual cluster means with RMSE below 0.3%,

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

confirming the decoder's utility as a mechanism for translating latent representations back into interpretable census variable profiles.

6.2 | Non-Linear Structure in Neighborhood Data

The systematic outperformance of autoencoders across all tested configurations provides empirical support for their use as a method to represent meaningful non-linear structure within neighborhood-level census data. PCA's reconstruction error at any given dimensionality represents the optimal linear approximation of the data; the autoencoder's ability to improve upon this bound therefore reflects variance that is attributable to non-linear relationships among the input variables. The magnitude of this non-linear information gain is not uniform across geographic contexts. The analysis revealed that it is greatest in areas with complex socio-demographic compositions: deprived urban neighborhoods, areas with high population density, and communities characterized by diverse ethnic composition, transient populations, or the co-occurrence of high educational attainment with economic disadvantage. These are precisely the contexts in which one would expect threshold effects, interaction terms, and non-stationary relationships to be most pronounced.

The geographic analysis of reconstruction errors also identified specific spatial contexts where linear methods exhibit systematic weakness. Regions with geographically concentrated cultural or linguistic identities, such as Cornwall and North Wales, produced elevated PCA reconstruction errors that the autoencoder substantially reduced. In these areas, variables relating to national identity have localized distributions that correlate weakly with other census variables at the national level. PCA, which constructs components to maximize variance across the full dataset, allocates limited representational capacity to such localized patterns. The autoencoder's non-linear encoding accommodates these configurations without requiring them to conform to globally linear relationships, resulting in more geographically equitable reconstruction performance. This finding carries implications for applied research: composite neighborhood measures that systematically underperform in specific geographic contexts risk introducing spatial bias into any downstream analysis or policy application that relies upon them.

6.3 | Contributions

This paper makes three principal contributions to the literature on composite measures of neighborhood structure and residential differentiation. First, it provides a detailed empirical evaluation of autoencoders as a dimensionality reduction method for census-based neighborhood analysis, benchmarked against PCA across a comprehensive range of latent dimensionalities using a national-scale dataset (188,880 Output Areas, 408 input variables). While prior work has applied autoencoders to geographic data in specific domains, including satellite imagery (Singleton et al. 2022), geodemographic preprocessing (De Sabbata and Liu 2023), and geographic embeddings (Cepeda et al. 2023; Agarwal et al. 2024), the present study offers a systematic comparison of linear and non-linear reduction for the specific purpose of deriving composite indicators of residential differentiation from census data. The evaluation extends beyond aggregate accuracy to examine the spatial distribution of reconstruction errors across deprivation gradients, density gradients, and geodemographic typologies, providing a granular account of where and why non-linear methods offer improvements.

Second, the study demonstrates that the practical advantages of autoencoder architectures, identified in Section 1.2 as addressing key constraints of non-parametric manifold methods, are realized in this application. The parametric encoder enables out-of-sample projection of new areas without recomputation of existing embeddings, an essential requirement for periodically updated indicators. The out-of-sample evaluation in Section 3.1 demonstrates this in practice, with areas withheld from training projected into the existing latent space and reconstructed without refitting the model. The decoder enables reconstruction-based diagnostics that quantify information preservation and loss at the level of individual variables and individual areas, and provides a principled mechanism for translating latent cluster centroids into interpretable census variable profiles. These capabilities directly address concerns regarding transparency and updatability that have limited the adoption of non-linear methods in operational neighborhood analysis.

Third, the case study application demonstrates that autoencoder-derived latent representations can serve as effective inputs for geodemographic classification, producing neighborhood typologies with strong spatial coherence and meaningful correspondence to established classification frameworks. Alongside the decoder's capacity to reconstruct interpretable cluster profiles, these results suggest that non-linear dimensionality reduction can provide a useful basis for census-based neighborhood classification, particularly where the aim is to combine dimensionality reduction, classification, and profile interpretation within a single modeling framework. Comparison with a PCA-based classification on the same inputs shows that approximately one in six OAs receives a different label depending on the representation used, with divergence concentrated in suburban, peri-urban, and older-population areas. This should not be read as establishing that one classification is correct and the other is not, since no ground-truth geodemographic partition exists. It does, however, illustrate that the choice of dimensionality reduction is not neutral for policy applications such as service targeting and resource allocation. Open and reproducible pipelines, such as that presented in this paper, allow that choice to be tested against the requirements of a given classification.

6.4 | Limitations

Several limitations should be acknowledged. The interpretability of autoencoder features remains less direct than that of principal components. Principal components can be examined through variable loadings that express each component as an explicit linear combination of the original inputs, enabling straightforward identification of the variables that drive each dimension of variation. Autoencoder features, by contrast, represent complex non-linear transformations for which no equivalent

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

closed-form expression exists. Although the decoder provides a mechanism for relating latent dimensions to original variables through reconstruction, and surrogate modeling approaches can approximate these relationships, the resulting interpretations are inherently less transparent than PCA loadings. For policy applications where stakeholders require clear and auditable explanations of how composite measures are constructed, this reduced transparency presents a genuine constraint.

Finally, the computational requirements of autoencoder training, while modest in absolute terms (under 10 min per model on GPU hardware, with a storage footprint under 2 MB), nonetheless exceed those of PCA, which admits a closed-form solution. For organizations without access to GPU infrastructure or deep learning expertise, this represents a practical barrier. Hyperparameter selection, including network depth, layer widths, activation functions, learning rate schedules, and regularization strength, introduces additional complexity that requires domain knowledge and systematic experimentation. Although the configurations tested in this study exhibited low sensitivity to random initialization, the broader hyperparameter space was not exhaustively explored, and optimal configurations may differ for datasets with substantially different characteristics.

6.5 | Future Directions

The findings of this study suggest several directions for future work. First, advances in more interpretable representation learning, including disentangled autoencoders and attention-based architectures, could be explored as a means of combining the non-linear modeling capacity demonstrated here with greater transparency of the learned features. Disentangled representations, in which individual latent dimensions correspond to distinct and separable data-generating factors, would be particularly valuable for policy applications that require each dimension of a composite measure to carry a clear substantive interpretation. Second, temporal extension presents an important opportunity. The parametric nature of the autoencoder means that models trained on one census period could, in principle, be applied to project data from subsequent periods into the same latent space, enabling analysis of neighborhood change within a consistent representational framework. Evaluating the stability and transferability of learned representations across census periods would be a valuable contribution to understanding residential dynamics. Third, the autoencoder framework could be extended to incorporate auxiliary objectives beyond reconstruction, such as downstream prediction targets, to produce latent representations that are jointly optimized for both representational fidelity and task-specific utility. Finally, the architecture evaluated here is aspatial, treating each Output Area independently of its neighbors. This is by design, since spatial encoding would change the question from whether non-linear autoencoders better capture census structure than PCA to whether spatial autoencoders outperform aspatial ones. Spatial structure could be introduced through the objective, via a spatial contiguity penalty on the latent representations, or through the architecture itself, with graph neural network components allowing the encoding to draw on the attributes of surrounding areas (De Sabbata and Liu 2023).

6.6 | Concluding Remarks

This study has established autoencoders as an effective and practical alternative to PCA for deriving composite measures of neighborhood structure from census data. The consistent improvements in reconstruction accuracy, the reduced spatial autocorrelation of residuals, and the more equitable distribution of errors across socio-economic and geographic gradients collectively demonstrate that non-linear dimensionality reduction captures information about residential differentiation that linear methods do not. The parametric nature of the encoder, the diagnostic capacity of the decoder, and the demonstrated utility of the latent space for geodemographic classification address several operational requirements that have constrained the adoption of non-linear methods in applied neighborhood analysis. While challenges of interpretability and computational accessibility remain, the evidence presented here supports the incorporation of autoencoder-based approaches into methods for constructing composite indicators of residential differentiation.

Funding

This paper was produced as part of the following ESRC grants: ES/L011840/1 and ES/Z50273X/1.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Code to reproduce all results in this paper, together with instructions for acquiring the publicly available census input data, is available at https://github.com/GeographicDataService/census_dimreduction_autoencoder. Census inputs were downloaded from Nomis (https://www.nomisweb.co.uk/sources/census_2021_bulk).

Endnotes

[1]https://data.geods.ac.uk/dataset/index-of-multiple-deprivation-imd.

[2]https://data.geods.ac.uk/dataset/output-area-classification-2021.

References

Agarwal, M., M. Sun, C. Kamath, et al. 2024. "General Geospatial Inference With a Population Dynamics Foundation Model." arXiv. https://doi.org/10.48550/arXiv.2411.07207.

Al-Shawamreh, S., and Y. Farhan. 2018. "GIS-Based Factorial Ecology and Social Public Space of the Twin City of Ramallah and Al-Bireh, West Bank, The Palestinian Authority." *Journal of Geographic Information System* 10, no. 3: 261–282. https://doi.org/10.4236/jgis.2018.103014.

Baldi, P., and K. Hornik. 1989. "Neural Networks and Principal Component Analysis: Learning From Examples Without Local Minima." *Neural Networks* 2, no. 1: 53–58. https://doi.org/10.1016/0893-6080(89)90014-2.

Barter, E., and T. Gross. 2019. "Manifold Cities: Social Variables of Urban Areas in the UK." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 475, no. 2221: 20180615. https://doi.org/10.1098/rspa.2018.0615.

Bassett, K., and J. R. Short. 1980. *Housing and Residential Structure: Alternative Approaches*. Routledge.

Batey, P. 2022. "Geodemographics and Urban Planning Analysis: An Historical Review." In *Theory and History in Regional Perspective: Essays in*

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Honor of Yasuhiro Sakai, edited by M. Kawano, K. Kourtit, P. Nijkamp, and Y. Higano, 3–39. Springer Nature (New Frontiers in Regional Science: Asian Perspectives). https://doi.org/10.1007/978-981-16-6695-7_1.

Belwalkar, A., A. Nath, and O. Dikshit. 2018. "Spectral-Spatial Classification of Hyperspectral Remote Sensing Images Using Variational Autoencoder and Convolution Neural Network." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-5: 613–620. https://doi.org/10.5194/isprs-archives-XLII-5-613-2018.

Berry, B. J. L., and J. D. Kasarda. 1977. *Contemporary Urban Ecology*. Macmillan.

Cepeda, V. V., G. K. Nayak, and M. Shah. 2023. "GeoCLIP: Clip-Inspired Alignment Between Locations and Images for Effective Worldwide Geo-localization." arXiv. https://doi.org/10.48550/arXiv.2309.16020.

Chen, X., C. Kitchen, and H. Kharrazi. 2025. "Assessing the Advantages and Disadvantages of Dimensionality Reduction Methods in Summarizing Housing Determinants of Health in the United States." *JAMIA Open* 8, no. 4: ooaf093. https://doi.org/10.1093/jamiaopen/ooaf093.

Coifman, R. R., and S. Lafon. 2006. "Diffusion Maps." *Applied and Computational Harmonic Analysis* 21, no. 1: 5–30. https://doi.org/10.1016/j.acha.2006.04.006.

Coifman, R. R., S. Lafon, A. B. Lee, et al. 2005. "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps." *Proceedings of the National Academy of Sciences of the United States of America* 102, no. 21: 7426–7431. https://doi.org/10.1073/pnas.0500334102.

Crawford, M. M., L. Ma, and W. Kim. 2011. "Exploring Nonlinear Manifold Learning for Classification of Hyperspectral Data." In *Optical Remote Sensing*, edited by S. Prasad, L. M. Bruce, and J. Chanussot, 207–234. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14212-3_11.

Da Vies, W. K. D. 1978. "Alternative Factorial Solutions and Urban Social Structure: A Data Analysis Exploration of Calgary in 1971." *Canadian Geographies/Géographies Canadiennes* 22, no. 4: 273–297. https://doi.org/10.1111/j.1541-0064.1978.tb01523.x.

De Sabbata, S., and P. Liu. 2023. "A Graph Neural Network Framework for Spatial Geodemographic Classification." *International Journal of Geographical Information Science* 37, no. 12: 2464–2486. https://doi.org/10.1080/13658816.2023.2254382.

Demšar, U., P. Harris, C. Brunsdon, A. S. Fotheringham, and S. McLoone. 2013. "Principal Component Analysis on Spatial Data: An Overview." *Annals of the Association of American Geographers* 103, no. 1: 106–128. https://doi.org/10.1080/00045608.2012.689236.

Diaz-Papkovich, A., L. Anderson-Trocmé, C. Ben-Eghan, and S. Gravel. 2019. "UMAP Reveals Cryptic Population Structure and Phenotype Heterogeneity in Large Genomic Cohorts." *PLoS Genetics* 15, no. 11: e1008432. https://doi.org/10.1371/journal.pgen.1008432.

Diaz-Papkovich, A., L. Anderson-Trocmé, and S. Gravel. 2021. "A Review of UMAP in Population Genetics." *Journal of Human Genetics* 66, no. 1: 85–91. https://doi.org/10.1038/s10038-020-00851-4.

Ericksen, E. G. 1949. "Review of the Social Areas of Los Angeles: Analysis and Typology." *American Sociological Review* 14, no. 5: 699. https://doi.org/10.2307/2086667.

Hale, R., and D. M. Austin. 1997. "An Exploratory Factor Model of Social Area Analysis." *Sociological Spectrum* 17, no. 1: 115–125. https://doi.org/10.1080/02732173.1997.9982154.

Harris, R., P. Sleight, and R. Webber. 2005. *Geodemographics, GIS and Neighbourhood Targeting*. Wiley.

Hawley, A. H., and O. D. Duncan. 1957. "Social Area Analysis: A Critical Appraisal." *Land Economics* 33, no. 4: 337–345. https://doi.org/10.2307/3144311.

Healy, J., and L. McInnes. 2024. "Uniform Manifold Approximation and Projection." *Nature Reviews Methods Primers* 4, no. 1: 82. https://doi.org/10.1038/s43586-024-00363-x.

Hirokane, S., and M. Amemiya. 2023. "Describing the Social Changing Pattern of Tsukuba Science City Using Social Area Analysis: Focusing on the Decade From 2010 to 2020." *Reports of the City Planning Institute of Japan* 22, no. 2: 268–274. https://doi.org/10.11361/reportscpij.22.2_268.

Hughes, J. W., and G. W. Carey. 1972. "Factorial Ecologies: Oblique and Orthogonal Solutions. A Case Study of the New York SMSA." *Environment and Planning A: Economy and Space* 4, no. 2: 147–162. https://doi.org/10.1068/a040147.

Hunter, A. A. 1972. "Factorial Ecology: A Critique and Some Suggestions." *Demography* 9, no. 1: 107–117. https://doi.org/10.2307/2060548.

Johnston, R. J. 1977. "Principal Components Analysis and Factor Analysis in Geographical Research: Some Problems and Issues." *South African Geographical Journal* 59, no. 1: 30–44. https://doi.org/10.1080/03736245.1977.9713493.

Jolliffe, I. T. 2002. *Principal Component Analysis*. 2nd ed. Springer.

Khan, W., S. Walker, and W. Zeiler. 2023. "A Bottom-Up Framework for Analysing City-Scale Energy Data Using High Dimension Reduction Techniques." *Sustainable Cities and Society* 89: 104323. https://doi.org/10.1016/j.scs.2022.104323.

Kingma, D. P., and M. Welling. 2019. "An Introduction to Variational Autoencoders." *Foundations and Trends in Machine Learning* 12, no. 4: 307–392. https://doi.org/10.1561/2200000056.

Klemmer, K., E. Rolf, C. Robinson, L. Mackey, and M. Rußwurm. 2024. "SatCLIP: Global, General-Purpose Location Embeddings With Satellite Imagery." arXiv. https://doi.org/10.48550/arXiv.2311.17179.

Liu, H., J. Yang, M. Ye, et al. 2021. "Using t-Distributed Stochastic Neighbor Embedding (t-SNE) for Cluster Analysis and Spatial Zone Delineation of Groundwater Geochemistry Data." *Journal of Hydrology* 597: 126146. https://doi.org/10.1016/j.jhydrol.2021.126146.

Liu, P., and S. De Sabbata. 2019. "Learning Digital Geographies Through a Multi-Modal Autoencoder." In *GISRUK Proceedings. GISRUK 2019, the 27th Annual GIScience Research UK Conference*. GISRUK.

Liu, Y., A. Singleton, and D. Arribas-Bel. 2019. "A Principal Component Analysis (PCA)-Based Framework for Automated Variable Selection in Geodemographic Classification." *Geo-Spatial Information Science* 22, no. 4: 251–264. https://doi.org/10.1080/10095020.2019.1621549.

Montosa Muñoz, J. C. 2023. "Application of the Method of Social Areas and Multivariate Analysis in Peri-Urban Areas of Andalusia (Spain)." *MOJ Ecology & Environmental Sciences* 8, no. 3: 117–127. https://doi.org/10.15406/mojes.2023.08.00281.

Rees, P. 1971. "Problems of Classifying Subareas Within Cities." In *City Classification Handbook: Methods and Applications*, edited by B. J. L. Berry, 265–330. Wiley-Interscience.

Rezende, D. J., S. Mohamed, and D. Wierstra. 2014. "Stochastic Backpropagation and Approximate Inference in Deep Generative Models." In *Proceedings of the 31st International Conference on International Conference on Machine Learning (ICML'14)*, vol. 32, II-1278–II-1286. JMLR.org.

Sainburg, T., L. McInnes, and T. Q. Gentner. 2021. "Parametric UMAP Embeddings for Representation and Semisupervised Learning." *Neural Computation* 33 (11): 2881-2907. https://doi.org/10.1162/neco_a_01434.

Salvati, L., M. Carlucci, and P. Serra. 2018. "Unraveling Latent Dimensions of the Urban Mosaic: A Multi-Criteria Spatial Approach to Metropolitan Transformations." *Environment and Planning A: Economy and Space* 50, no. 1: 93–110. https://doi.org/10.1177/0308518X17736313.

Schölkopf, B., A. Smola, and K.-R. Müller. 1998. "Nonlinear Component Analysis as a Kernel Eigenvalue Problem." *Neural Computation* 10, no. 5: 1299–1319. https://doi.org/10.1162/089976698300017467.

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Shevky, E., and W. Bell. 1955. *Social Area Analysis*. Stanford University Press.

Shevky, E., and M. Williams. 1949. *The Social Areas of Los Angeles: Analysis and Typology*. University of California Press.

Singleton, A., D. Arribas-Bel, J. Murray, and M. Fleischmann. 2022. "Estimating Generalized Measures of Local Neighbourhood Context From Multispectral Satellite Images Using a Convolutional Neural Network." *Computers, Environment and Urban Systems* 95: 101802. https://doi.org/10.1016/j.compenvurbsys.2022.101802.

Singleton, A., and S. E. Spielman. 2014. "The Past, Present, and Future of Geodemographic Research in the United States and United Kingdom." *Professional Geographer* 66, no. 4: 558–567. https://doi.org/10.1080/00330124.2013.848764.

Singleton, A., and S. E. Spielman. 2026. "Geodemographics and Residential Differentiation: A Methodological Review and Future Directions for Learned Representations of the Social Landscape." *Computers, Environment and Urban Systems* 125: 102396. https://doi.org/10.1016/j.compenvurbsys.2025.102396.

Sousa, D., and C. Small. 2022. "Joint Characterization of Spatiotemporal Data Manifolds." *Frontiers in Remote Sensing* 3: 760650. https://doi.org/10.3389/frsen.2022.760650.

Timms, D. W. G. 1971. *The Urban Mosaic: Towards a Theory of Residential Differentiation*. Cambridge University Press.

van der Maaten, L., and G. Hinton. 2008. "Visualizing Data Using t-SNE." *Journal of Machine Learning Research* 9: 2579–2605.

van der Maaten, L. 2009. "Learning a Parametric Embedding by Preserving Local Structure." In *International Conference on Artificial Intelligence and Statistics*. PMLR. https://api.semanticscholar.org/CorpusID:18203584.

Wang, Y., H. Yao, and S. Zhao. 2016. "Auto-Encoder Based Dimensionality Reduction." *Neurocomputing* 184: 232–242. https://doi.org/10.1016/j.neucom.2015.08.104.

Williams, C. K. I., and M. W. Seeger. 2000. "Using the Nyström Method to Speed Up Kernel Machines." *Neural Information Processing Systems* 13: 682–688.

Wyszomierski, J., P. A. Longley, A. D. Singleton, C. Gale, and O. O'Brien. 2023. "A Neighbourhood Output Area Classification From the 2021 and 2022 UK Censuses." *Geographical Journal* 190, no. 2: e12550. https://doi.org/10.1111/geoj.12550.

Xiu, G., and H. Chen. 2023. "Unravelling the Variations of the Society of England and Wales Through Diffusion Mapping Analysis of Census 2011." *Journal of the Royal Society Interface* 20, no. 205: 20230081. https://doi.org/10.1098/rsif.2023.0081.

15384632, 2026, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/gean.70059 by NICE, National Institute for Health and Care Excellence, Wiley Online Library on [08/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License