applied sciences

MDPI

Article

A Comparative Analysis of Distance Metrics for Multilabel Land-Use Classification Using Mobile Phone Data in Trento

Manuel Mendoza-Hurtado [1,*], Jose Manuel Cuevas-Muñoz [2] and Domingo Ortiz-Boyer [2,*]

[1] Department of Computer Science, University of Jaen, Linares Campus, D Building D-151, 23700 Linares, Spain
[2] Department of Computer Science and Artificial Intelligence, University of Cordoba, Campus de Rabanales, Albert Einstein Building, 14071 Cordoba, Spain; i62cumuj@uco.es
* Correspondence: mmhurtad@ujaen.es (M.M.-H.); dortiz@uco.es (D.O.-B.)

Featured Application

A comparison of distance functions for exploratory multilabel land-use inference from aggregated mobile-network activity in heterogeneous urban and regional settings.

Abstract

Mobile-network traces provide a complementary view of land use, yet the role of distance functions in multilabel neighborhood-based inference remains underexplored in heterogeneous regions. We compare eleven pointwise, complexity-invariant, and elastic distances across four multilabel *k*-nearest-neighbor (*k*-NN) algorithms using weekly profiles of incoming calls, outgoing calls, and internet activity from 241 manually labeled cells in Trentino, Italy. Four non-*k*-NN classifiers with fixed settings constitute the reference baselines. We evaluate performance through nine different metrics under the same iterative stratified 10-fold cross-validation procedure. The distances belonging to the Canberra family perform well in several classifier–channel pairs, including an average precision of 0.905 and a ranking loss of 0.125 in ML-localkNN for incoming calls. The experiment also highlights the inconsistent effects of complexity correction and varying sensitivities to distance selection across classifiers and channels. The study contributes a systematic empirical assessment of distance families, a consistent multilabel evaluation framework, and spatial diagnostics linking numerical performance to land-use interpretation.

Keywords: call detail records; digital traces; distance metrics; k-nearest neighbors; land-use classification; mobile phone data; multilabel classification; Trento; urban computing; urban data science

Check for updates

Academic Editor: Salvador García-Ayllón Veintimilla

Received: 1 October 2026
Revised: 3 November 2026
Accepted: 5 November 2026
Published: 7 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Inferring how land is used across a region is a central problem in urban data science and a prerequisite for sustainable planning, environmental monitoring, and mobility management. Traditional surveys provide authoritative land-use maps but are expensive to maintain and quickly become outdated in rapidly evolving peri-urban areas. Mobile-network data, including call detail records (CDRs), offer a complementary source: the aggregate volume and temporal rhythm of calls, messages, and data sessions recorded at each base station or grid cell encode a distinctive "activity signature" that correlates with the functional mix of an area [1–4]. When these digital traces are released in spatially gridded, anonymized form, as in the Telecom Italia Big Data Challenge [5], they

enable reproducible benchmarking of land-use inference methods across regions and telecommunications channels.

A common limitation of past research is the assumption of a single land-use class for each spatial cell. In reality, a cell may combine residential, business, and recreational functions, especially in regions that extend from compact urban centers to rural settlements and forest areas [6,7]. Multilabel learning allows such functional mixtures to be modeled explicitly. Instance-based learning methods, specifically *k*-nearest-neighbor approaches and their variations, are a natural choice, since they depend on similarities between training and query instances [8,9].

In all *k*-NN approaches, similarity is defined by the distance function. The application of Euclidean distance is standard practice in many cases; however, its appropriateness should be tested for CDR-derived activity signatures, whose rhythm is very pronounced both on diurnal and weekly timescales but with absolute amplitudes differing by orders of magnitude from one cell to another. Alternative distance functions, like ratio-based distances scaled locally [10], feature-bounded distances [11], or complexity-compensated composite distances [12], provide different methods for comparing weekly rhythms and taking into account differences in the volume of activity. Still, there is very limited empirical evidence on their performance in multilabel land-use inference [13]. The time-series classification literature has demonstrated that the distance metric is an important modeling decision [14,15]. Such a conclusion has not been tested yet in the context of multilabel urban function inference.

Trentino is a diverse testing ground for this investigation. The selected study area includes the urban core of Trento, suburban settlements, and montane forests. The label set includes an explicit *forest* category, whose activity profiles tend to be flat and low in amplitude. In this setting, ratio-based metrics may behave differently from magnitude-sensitive distances. Whether this difference improves classification is treated as an empirical question rather than an assumed property.

The present study addresses this gap by conducting an empirical analysis of 11 distance functions across four multilabel *k*-NN classifiers applied to three CDR-based activity channels in the Trentino grid. The main contributions are as follows:

1. **Heterogeneous regional landscape distance metric evaluation:** We compare five pointwise base metrics, five complexity-invariant composite metrics, and one elastic (DTW-based) distance in multilabel *k*-NN pipelines. We test these pipelines in an area combining urban, suburban, and forested zones. This extends the scope of distance-metric studies beyond metropolitan settings.
2. **Comparison across classifiers and activity channels:** We evaluate four multilabel *k*-NN classifiers (BR-kNN, ML-kNN, LAML-kNN, and ML-localkNN) alongside four non-*k*-NN reference models with fixed configurations. Previously selected, metric-specific neighborhood configurations are retained and reevaluated using consistent metric definitions.
3. **Comprehensive evaluation and spatial interpretation:** Nine complementary measures and a unified evaluation protocol enable consistent comparisons of classification and label ranking performance. Maps of spatial predictions link this quantitative assessment to land use interpretation and delineate regions of transition for further analysis.

From the viewpoint of urban data science, the current paper introduces distance selection as an explicit and interpretable part of multilabel land-use inference. Our empirical comparisons show how the performance of distance functions varies across classifiers and activity channels. This forms a foundation for the selection of potential distances for further regional assessments. The remainder of this article is organized as follows: Section 2

reviews related work on CDR-based land-use inference, multilabel classification, and distance metrics. Section 3 defines the 11 distance functions and discusses their relevance to activity signatures. Section 4 describes the Trentino dataset and the experimental protocol. Section 5 reports predictive performance under the common evaluation framework and presents qualitative spatial visualizations. Section 6 interprets the findings, and Section 7 summarizes the study.

2. Related Work

Inferring functional land use from digital traces is an active area of urban computing and urban data science. This is partly because traditional land-use surveys are costly and updated infrequently [16]. Early studies have shown that aggregated mobile-network traces contain temporal signatures, such as daytime peaks, overnight troughs, and differences between weekdays and weekends, which can inform citywide land-use classification. Soto and Frías-Martínez demonstrated that activity patterns inferred from mobile phone traces can generate land-use signatures and that these patterns can be transferred across cities at coarse spatial resolutions [17]. Toole et al. formalized how time-dependent activity profiles represent "actual" land use and reveal differences from officially designated land use [1]. Pei et al. quantified the influence of base-station density and land-use heterogeneity on classification difficulty [2], while Louail et al. used aggregated CDR data to reveal the spatial organization and temporal rhythms of activity hotspots [4]. Broader surveys have since established CDR-based analysis as a research area in its own right [18].

Open, gridded telecommunications data have played an important role in reproducible experimentation. The Telecom Italia Big Data Challenge dataset, described by Barlacchi et al., provides activity grids for multiple channels in Milan and Trentino at regular spatial and temporal intervals [5]. This resource has also been used to relate urban vitality to land-use diversity [19]. These datasets present several modeling challenges: activity volumes are highly skewed, spatial autocorrelation is common, and many cells have multiple functions that cannot be represented by a single label.

2.1. Multilabel Representations of Functional Mixtures

Traditional methods of land use classification assign each spatial unit to a single category based on zoning rules. Yuan et al. suggested identifying functional zones by combining movement traces and points of interest [20]; yet, similar to many early proposals, their approach involves assigning mutually exclusive categories. However, individual spatial units might support multiple functions (a cell may house apartments, shops, and green spaces), and the main functionality might change during the day. Multilabel learning is able to handle this functional diversity by allowing each spatial unit to be associated with some subset of a predefined label set. According to the SAMPLID framework, clustering-based pipelines are prone to optimize geometric criteria that might not align well with semantic land-use classifications; hence, it suggests considering supervised alternatives based on task-specific ground truth [21]. Extending this reasoning, MAPLID considers place identification as a supervised multilabel classification task using CDR-based features [6]; the methodology was later applied to Trentino [7]. Related research on home and work location detection highlights the need to validate mobility-derived functional proxies using predictive models when reference labels are available [22,23].

2.2. Instance-Based Multilabel Learners

The multilabel classification (MLC) literature encompasses problem-transformation strategies (e.g., Label Powerset and RAkEL ensembles), algorithm-adaptation methods (e.g., ML-kNN), and end-to-end deep learning architectures. Extensive benchmarks show that

the relative ranking of methods depends on dataset characteristics, label cardinality, class imbalance, and the evaluation metrics used [15,24]. Among instance-based approaches, ML-kNN estimates per-label posterior probabilities from label frequency counts among the k nearest neighbors [8]. Because those neighbors are selected using pairwise distances, changing the distance function can change the posterior estimates, making the predictions sensitive to distance choice, particularly when label distributions vary across the study area.

To address this sensitivity, previous studies have proposed some improvements in neighborhood design. Local adaptation is one of them, and it suggests using different values of k for different parts of the feature space, since one global value of k may not be optimal due to the different complexities of multilabel decision boundaries [25]. Label-specific adaptations further tune the neighborhood size based on the label frequencies and separabilities [26]. Metric learning is another technique that optimizes the distance function itself with respect to multilabel k-NN objectives [27], and it involves using label-specific local metrics or similarity functions that account for multiple semantic relations between labels [28,29]. In contrast, the present study focuses on predefined analytical distance functions.

2.3. Distance Functions for Activity-Derived Time Series

Weekly activity profiles can be viewed as short, fixed-length time series where pairwise distances define which spatial entities are considered functionally similar. There is much research on time series classification indicating that the selection between pointwise, elastic, and feature-based distances can greatly impact accuracy and efficiency [14,30,31]. In multivariate cases, elastic methods do not always outperform simpler ones when both are properly tuned, and preprocessing, especially normalization, can have as much effect as the distance itself [32,33]. Dynamic time warping (DTW), a popular elastic distance, allows for phase shifting due to nonlinear alignment of sequences; there are differentiable versions like soft-DTW which allow for learning with gradient-based optimization in the DTW geometry [34]. However, in commuting-related activity profiles, peak timing may carry functional information, making pointwise distances relevant alternatives to elastic methods. Pointwise distances thus remain reasonable options to compare empirically.

Complexity-invariant corrections form a related family of methods. The complexity-invariant distance (CID) rescales a base distance by the ratio of the two series' complexities to account for differences in local variability [12]. Its usefulness for land-use inference depends on whether variability within a profile reflects noise or distinct functional patterns. In aggregated CDR data, variability often correlates with the sharpness of commuting peaks. The value of the correction must therefore be assessed empirically, particularly with multilabel k-NN methods, whose predictions can be sensitive to small changes in neighborhood composition.

2.4. Validation Under Spatial Dependence

Gridded spatial data call into question the independence assumption behind standard cross-validation, because neighboring cells typically share infrastructure, demographic characteristics, and mobility flows. Even iterative stratification, which preserves multilabel proportions across folds [35,36], does not guarantee spatial separation, and it may give optimistic performance estimates when adjacent cells fall in different folds. Geospatial machine-learning studies therefore advocate blocked or buffered cross-validation when the goal is geographic generalization [37,38]. We use iterative stratified k-fold cross-validation for multilabel benchmarking. Complementary spatially explicit validation would strengthen conclusions about transferability beyond the study area [39].

Taken together, the literature indicates that (a) aggregated mobile-network activity patterns encode land-use information [1,2,17], (b) multilabel formulations can represent functional mixtures in supervised pipelines [6,21], (c) multilabel k-NN methods provide interpretable baselines whose outputs depend on the distance function [8,25,26], and (d) distance choice is an important modeling decision in the classification of temporal activity signatures [14,31]. However, distance families have not been systematically evaluated for multilabel k-NN land-use inference in a geographically heterogeneous setting that includes forested and low-density areas alongside urban areas. This gap motivates the present benchmark for urban data science pipelines that transform mobile phone traces into interpretable land-use maps.

3. Distance Metrics for Land-Use Signatures

3.1. Characteristics of CDR-Based Activity Signatures

The structural properties of CDR-derived features motivate the comparison of distance functions. In the four k-NN classifiers considered here, distances determine the neighborhoods from which label predictions or posterior probabilities are derived. Distance choice therefore directly links the activity representation to the inferred functional labels.

Three properties of the Trentino activity data are of particular relevance. **(i) Clock-time alignment.** Each cell is summarized by a 168-dimensional vector (24 h $\times$ 7 days) formed by binning raw event counts into hourly slots. These profiles retain daily and weekday–weekend rhythms on a common clock-time grid. The timing of activity peaks can convey functional information, for example, when residential and commercial areas differ in their morning or midday activity. Pointwise distances maintain these timing contrasts, whereas elastic distances allow some temporal displacement. Comparing both constructions therefore addresses a meaningful modeling choice rather than assuming that alignment is always beneficial.

(ii) Amplitude heterogeneity. The activity levels exhibited by different urban cores vary substantially compared to those of peripheral or montane cells. Euclidean and Manhattan distances prioritize the absolute differences of the components, while the Canberra distance scales the difference by the sum of the components. Euclidean and Manhattan distances are therefore more sensitive to absolute differences in activity volume, while Canberra emphasizes proportional differences, particularly in low-activity intervals. A comparison of these approaches is especially relevant when the study area includes both high-activity urban centers and low-activity peripheral cells.

(iii) Functionally informative variability. The CID correction [12] uses the complexity of two time series to rescale a base distance metric. Variability in CDR-based profiles could be a result of the measurement process or the underlying activity patterns, such as commuting or continuous daytime activity. Comparison of the base distance and the corresponding CID distance gives us an empirical measure of the usefulness of this rescaling.

Together, clock-time alignment, amplitude heterogeneity, and functionally informative variability motivate the three distance families evaluated below. They also provide an interpretive framework for discussing the observed classifier–channel differences.

3.2. Base Metrics

The 11 distance functions fall into three groups: (i) five pointwise base metrics that compare vectors element by element, (ii) five composites that multiply each base metric by a CID correction factor, and (iii) one elastic metric that combines constrained DTW with the CID correction. Throughout, $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ denote feature vectors of length n.

Euclidean. The standard L_2 norm:

$$d_{\text{Eucl}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}\,. \tag{1}$$

Squaring the component differences gives greater weight to large discrepancies, such as those associated with midday calling peaks in the Trentino urban core. Consequently, matching peak volumes may make two cells appear similar even when differences in their off-peak or weekend rhythms are informative about land use.

Manhattan. The L_1 norm:

$$d_{\text{Manh}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|\,. \tag{2}$$

Replacing squared differences with absolute differences decreases the impact of large differences as compared to the Euclidean measure. But the Manhattan distance does not normalize each component separately; hence, cells with similarly shaped profiles but different activity volumes can be far apart.

Logarithmic. A logarithmic-scale variant that compresses large differences:

$$d_{\text{Log}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \log\bigl(1 + |x_i - y_i|\bigr)\,. \tag{3}$$

The logarithmic transformation reduces the relative influence of large component-wise differences while preserving their contribution to the overall distance. Unlike ratio-based distances, it does not normalize differences by the magnitudes of the corresponding components. Instead, it provides a nonlinear transformation of absolute differences, reducing the dominance of large deviations without explicitly measuring proportional differences.

Hassanat (HasD). A dimensionality-invariant metric [11] that bounds each feature's contribution to $[0, 1]$:

$$d_{\text{HASD}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} D_i, \quad D_i = \begin{cases} 1 - \dfrac{1 + \min(x_i, y_i)}{1 + \max(x_i, y_i)} & \text{if } \min(x_i, y_i) \geq 0, \\ 1 - \dfrac{1 + \min(x_i, y_i) + |\min(x_i, y_i)|}{1 + \max(x_i, y_i) + |\min(x_i, y_i)|} & \text{otherwise.} \end{cases} \tag{4}$$

Bounding each component's contribution limits the influence of any single hourly bin. The added constant also ensures a nonzero denominator for nonnegative, low-activity components. Hassanat thus provides a bounded alternative whose classification performance can be compared with that of the other distances.

Canberra. First proposed by Lance and Williams [10], this ratio-based metric amplifies discrepancies where both operands are small:

$$d_{\text{CanD}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|}\,. \tag{5}$$

We define $0/0 = 0$, so an hourly slot in which both cells are inactive contributes no dissimilarity. Dividing the absolute difference by the sum of the component magnitudes emphasizes proportional rather than absolute variation. A small absolute difference between low-activity components can therefore contribute substantially to the total distance, making Canberra a relevant candidate for comparing heterogeneous weekly profiles.

3.3. Complexity-Invariant Composites (CID)

Batista et al. [12] proposed multiplying any base distance d_b by a correction factor that accounts for differences in the "complexity" (i.e., local variability) of the two series. Series complexity is quantified as:

$$CE(\mathbf{x}) = \sqrt{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}\,, \tag{6}$$

and the correction factor is the ratio of the larger to the smaller complexity:

$$CF(\mathbf{x}, \mathbf{y}) = \frac{\max\big(CE(\mathbf{x}),\, CE(\mathbf{y})\big)}{\min\big(CE(\mathbf{x}),\, CE(\mathbf{y})\big)}\,. \tag{7}$$

The composite CID distance is $d_{\text{CID}}(\mathbf{x}, \mathbf{y}) = d_b(\mathbf{x}, \mathbf{y}) \times CF(\mathbf{x}, \mathbf{y})$.

Applying this correction to the five base metrics produces five composites: **CID–Euclidean**, **CID–Manhattan**, **CID–Log**, **CID–Hassanat**, and **CID–Canberra**. Its usefulness in Trentino is uncertain (Section 3.1): differences in complexity may reflect measurement effects or functional contrasts, such as the pronounced morning and evening peaks of a commuting suburb compared with the nearly flat profile of a forest cell. Rescaling these differences may therefore alter neighborhood structure in ways that help or hinder classification.

3.4. DTW-Based Metric

CID-DTW10. We include one elastic distance: DTW [40] constrained by a Sakoe–Chiba band of width $w = 10$ and multiplied by the CID correction factor:

$$d_{\text{CID-DTW10}}(\mathbf{x}, \mathbf{y}) = \text{DTW}_{w=10}(\mathbf{x}, \mathbf{y}) \times CF(\mathbf{x}, \mathbf{y})\,. \tag{8}$$

Each weekly profile consisting of 168 elements is fed into the DTW algorithm as a single sequence. The band constraint excludes extreme warping paths while allowing limited realignments in time.

The CID-DTW10 algorithm is based on two model assumptions: band-constrained temporal warping and complexity correction. We selected just this elastic comparator from the benchmark set to make the comparison more focused; we did not evaluate the uncorrected DTW algorithm separately. The warping window restricts the range of the alignment, while the CID factor scales the distance obtained by the warping without increasing computational cost compared to DTW with the same window. One possible explanation of the CID-DTW10 results could be that the differences in timing of the activity peaks encode some land use information, which warping removes.

3.5. Computational Complexity

Each pointwise base metric makes a single pass over the $n = 168$ feature dimensions, giving a computational cost of $\mathcal{O}(n)$ per pair. The CID correction requires an additional $\mathcal{O}(n)$ pass to estimate the complexity of each series, so the five CID composites retain linear complexity. Unconstrained DTW requires up to $\mathcal{O}(n^2)$ alignment updates, whereas a Sakoe–Chiba band of width w reduces this to $\mathcal{O}(n \cdot w)$. Comparing a query cell with N training cells therefore requires $\mathcal{O}(N \cdot n^2)$ and $\mathcal{O}(N \cdot n \cdot w)$ updates, respectively. These nominal operation counts motivate the use of a constrained window.

4. Experimental Setup

4.1. Methodological Overview

The design assesses distance choice jointly with classifier choice and activity channel. Weekly profiles of incoming calls, outgoing calls, and internet activity from 241 labeled cells are evaluated with eleven distance functions and four multilabel k-NN classifiers. Each combination is assessed under iterative stratified 10-fold cross-validation, using nine complementary metrics and the metric-specific configurations described below. Four non-k-NN classifiers with fixed configurations provide reference baselines, and spatial prediction maps support interpretation of the numerical results. All of these components sit within one evaluation framework for land-use classification in heterogeneous urban and regional settings. Figure 1 summarizes the workflow and highlights three empirical contributions: extending distance assessment to an urban, suburban, and forest setting (C1); jointly examining distance functions, classifiers, and activity channels (C2); and combining aggregate evaluation, per-label diagnostics, and spatial interpretation (C3).

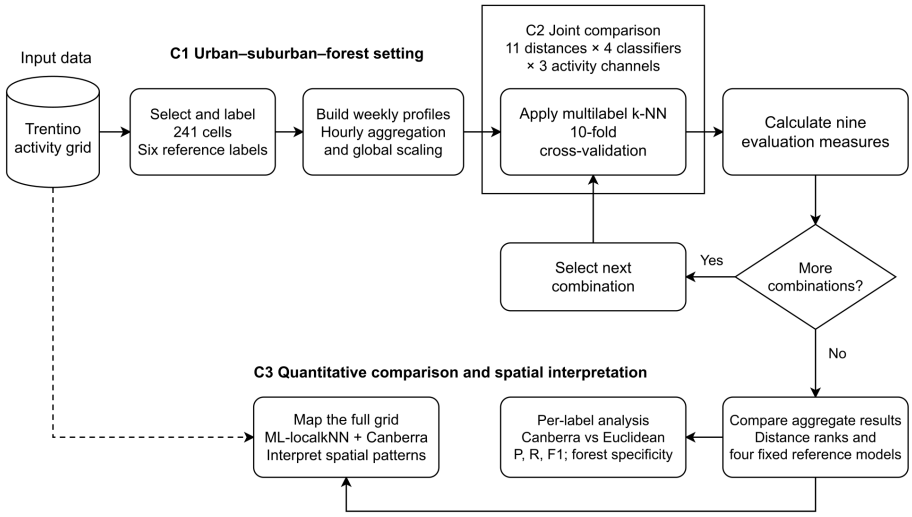


Figure 1. Methodological workflow of the Trentino distance metric benchmark. C1–C3 identify the empirical contributions described in the text. P and R denote precision and recall, respectively.

4.2. Dataset and Feature Construction

The experimental data come from the Telecom Italia Big Data Challenge [5], which records telecommunications activity at 10-minute intervals across a grid covering Trentino. The source dataset spans two months, from 1 November 2013 to 1 January 2014, and contains five channels: incoming calls, outgoing calls, incoming short message service (SMS) messages, outgoing SMS messages, and internet sessions. Following prior studies that identified the call and internet channels as the most discriminative for land-use tasks [6,7], we retain incoming calls (*callin*), outgoing calls (*callout*), and internet usage (*internet*) and exclude the two SMS channels.

The period chosen for the current analysis is 4–10 November 2013, a week-long span with no national holidays. This time frame allows us to cover the entire weekly cycle, including daily rhythms and weekday–weekend differences that are often detectable in CDRs [1,2], while making the comparison of different multilabel classifiers computationally feasible. Based on our previous experience with MAPLID in Milan and Trento, where weekly CDR profiles aggregated into hourly bins were found to be informative for multilabel place recognition [6], this representation was chosen. To the best of the authors' knowledge, there were no special city-wide events impacting the analyzed activity within this week.

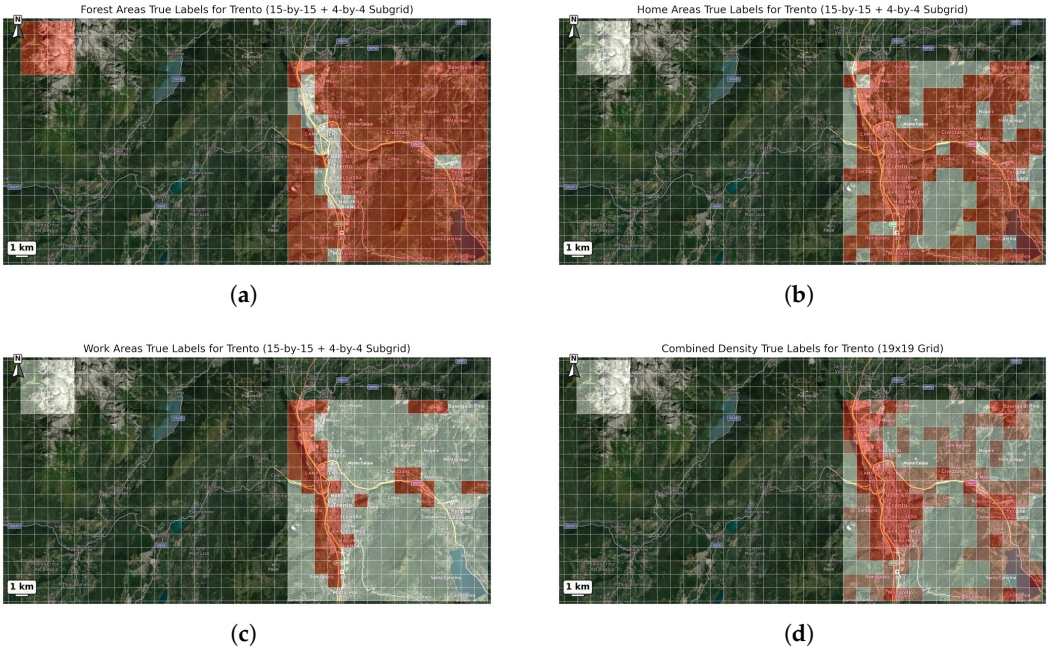
The Autonomous Province of Trento (Trentino) is an Alpine region characterized by settlements and economic activities that are confined to valleys and towns such as Trento and Rovereto. It is one of the wealthiest regions in Italy, with a GDP per capita above the national average in 2013 [41]. The presence of urban centers, dispersed settlements, and natural landscapes makes it an interesting case for analyzing heterogeneous land-use patterns.

The study area includes a 15-by-15 subgrid and a 4-by-4 forested subgrid inside the provincial grid, resulting in a total of 241 grid cells. It includes central Trento, suburban settlements, and montane forests and contains all six labels of the benchmark.

These two subgrids were annotated manually by domain experts during the creation of the dataset following a hierarchical three-step visual procedure. First, a label of morphological density (*high*, *medium*, or *low*) was assigned based on satellite images by estimating the building compactness and the visually assessed percentage of impervious surfaces: above 80%, 25–80%, and below 25%, respectively. These percentages are visual annotation thresholds and not independent measurements of the surface fractions. Second, labels of the function of buildings (*home*, *work*, or both) were determined based on their type and context. Large buildings of an industrial or office type were labeled as *work*, apartment buildings or subdivisions as *home*, and mixed-type blocks and campus housing next to offices as both. Third, ambiguous cases were checked using OpenStreetMap tags, Google Street View, and cartographic information available from the data portal of the Autonomous Province of Trento [42]. Disagreements between annotators were resolved through a joint review, prioritizing provincial cartographic information.

The additional, nonexclusive label *forest* was used if there was a presence of forest or natural vegetation in the cell, based on satellite imagery and land cover maps; the mere presence of vegetation was not enough. The archival data do not provide information on the date of imagery and the interrater agreement before the consensus. We therefore refer to the consensus annotations as *reference labels*, rather than verified ground truth.

The reference labels *forest*, *home*, and *work* are shown in Figure 2a–c, and the composite high-, medium-, and low-density labels are shown in Figure 2d.



(**a**) (**b**)

(**c**) (**d**)

Figure 2. Manually assigned reference labels used to interpret the Trento study area: (**a**) forest; (**b**) home; (**c**) work; and (**d**) combined density. Red represents the positive labels and white negative ones.

The raw 10-minute data were aggregated into hourly intervals, yielding 168 dimensions per feature vector (24 h × 7 days) and three multilabel datasets, one per channel. Before distance calculation, all hourly input features within each channel were normalized to [0, 1] using a single global scaling factor [43]. This factor was estimated from the complete study dataset, and normalization was applied before the data were partitioned into folds, ensuring that all configurations for a given channel used the same preprocessed inputs. The implications of this choice are discussed in Section 6.

The six labels are defined as follows:

- **Work**: predominantly commercial, office, or institutional land use.
- **Home**: predominantly residential land use.
- **Forest**: predominantly natural or forested land cover.
- **High density**: high built-up density.
- **Medium density**: intermediate built-up density.
- **Low density**: low built-up density.

A cell may have multiple labels. The three density labels are mutually exclusive, whereas *home*, *work*, and *forest* may co-occur with a density label. The observed prevalences are: home 155/241 (64.3%), work 44/241 (18.3%), forest 220/241 (91.3%), high density 49/241 (20.3%), medium density 89/241 (36.9%), and low density 103/241 (42.7%). The high prevalence of *forest*, in particular, means that aggregate scores cannot establish class-specific discrimination. Supplementary Table S2 provides per-label diagnostics for selected ML-localkNN configurations, including specificity for identifying the 21 non-forest cells.

Standard multilabel descriptors [44] characterize the selected study area as moderately multilabel: the average label cardinality is 2.739 (approximately three active labels per cell), the label density is 0.456, and 15 of the 64 theoretically possible label combinations (2^6) are represented.

4.3. Multilabel k-NN Classifiers

Four multilabel *k*-NN algorithms serve as the distance-sensitive core of the benchmark:

- **BR-kNN** [9]: decomposes the multilabel problem into one independent *k*-NN binary classifier per label and aggregates the individual decisions.
- **ML-kNN** [8]: applies Bayesian reasoning to the label frequency vector observed among the *k* closest training instances, producing per-label posterior probability estimates.
- **LAML-kNN** [45]: augments the *k*-NN neighborhood with local label-awareness mechanisms that adapt predictions to the label composition of each region.
- **ML-localkNN** [25]: extends ML-kNN with two kinds of locality, local distance weighting and label-specific neighborhood priors, which makes it the classifier most sensitive to the underlying distance geometry.

Including classifiers that use distance information in different ways allows us to assess whether the benefit of changing distance functions depends on how each classifier uses its neighborhood [25,26].

Four non-*k*-NN multilabel classifiers with fixed configurations provide reference baselines: Classifier Chains (**CC**) [46] with a support vector machine (SVM) base learner using a radial basis function (RBF) kernel; Label Powerset (**LP**) and Random *k*-Labelsets with disjoint partitions (**RAkEL-d**) [47], each using a depth-limited random forest (max_depth = 2); and Multilabel Conditional Random Fields (**ML-CRF**), trained using the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm for up to 100 iterations with $c_1 = c_2 = 0.1$ and unrestricted label transitions. RAkEL-d used a disjoint labelset size of 1, a conservative decomposition choice. These baselines were not exhaustively tuned and serve as fixed reference points for the *k*-NN benchmark. The study assesses distance effects within this benchmark and does not establish a ranking of classifier families.

4.4. Evaluation Protocol

4.4.1. Cross-Validation and Parameter Selection

All experiments use iterative stratified 10-fold cross-validation [35,36]. In iterative stratification, examples are allocated such that each fold represents the overall multilabel distribution, which makes it a better method for assessing performance compared to random stratification where multiple labels overlap.

Neighborhood settings were selected separately for each evaluation objective. For BR-kNN, ML-kNN, and LAML-kNN, the global neighborhood size was varied over $k \in \{1, \ldots, 25\}$. ML-localkNN used a local neighborhood size derived from the training data, with sorting and averaging settings selected for each objective. Settings selected using the original metric definitions were retained and reevaluated using the consistent definitions below, without retuning, to preserve comparability with the existing benchmark. For BR-kNN accuracy, k was selected over $k = 1, \ldots, 25$ using the mean Jaccard similarity across folds. The same ten folds are used for selection and reporting.

Tables in the next section represent a metric-specific empirical upper-bound benchmark under the historical selection scheme, rather than the performance of a single production model. Here, "upper-bound" describes a comparison conditioned on parameter selection; it is not a guaranteed mathematical bound or an independently validated estimate of generalization performance. The reuse of the folds for the selection and reporting of results can lead to overly optimistic performance evaluations [48]. The exact global k values and the ML-localkNN `sort_k`, `nmedia`, and objective codes are provided in Supplementary Table S1 for each classifier–metric pair. For ML-localkNN, the actual k differs per instance and fold; the historical average is presented separately.

4.4.2. Evaluation Metrics

Nine complementary metrics are used to quantify multilabel performance, employing a common set of evaluation definitions. Let Y_i and $\widehat{Y}_i$ denote the reference and predicted label sets for instance i, let $L = 6$ be the number of labels, and let $s_{i\ell}$ be the continuous score for label ℓ. Each metric is calculated within each test fold. The resulting tables in the next section report the unweighted means of the ten fold scores and the population standard deviation of these scores (ddof $= 0$), rather than a pooled score across all 241 cells. The separate per-label analysis below uses pooled out-of-fold predictions.

- **Accuracy** (↑): instance-averaged Jaccard similarity between predicted and reference label sets.
- **Hamming loss** (↓): mean $|Y_i \triangle \widehat{Y}_i|/L$, the fraction of incorrectly assigned label–instance pairs.
- **F1 score** (↑): mean of the instance values $F_i = 2|Y_i \cap \widehat{Y}_i|/(|Y_i| + |\widehat{Y}_i|)$.
- **ROC AUC** (↑): micro-averaged area under the receiver operating characteristic (ROC) curve, pooling scores and binary reference indicators for all label–instance pairs within a fold. Ties between positive and negative scores receive half credit. This metric is not the mean of the per-label AUCs.
- **Coverage** (↓): mean zero-based depth needed to retrieve all relevant labels. Tied scores receive their largest possible rank.
- **Ranking loss** (↓): mean fraction of relevant–irrelevant label pairs with $s_{i\ell} \leq s_{im}$, so a tie counts as an ordering error.
- **Average precision** (↑): mean noninterpolated average precision (AP), calculated separately for each instance with tied label scores treated as a single threshold.
- **Recall** (↑): mean of the instance values $R_i = |Y_i \cap \widehat{Y}_i|/|Y_i|$.
- **Precision** (↑): mean of the instance values $P_i = |Y_i \cap \widehat{Y}_i|/|\widehat{Y}_i|$.

An upward arrow means that higher values are better for the corresponding metric, while a downward arrow indicates that lower values are better for the corresponding metric.

The term accuracy refers to instance-based Jaccard similarity, $J_i = |Y_i \cap \widehat{Y}_i| / |Y_i \cup \widehat{Y}_i|$ (higher is better), rather than exact-match (subset) accuracy. When both the reference and predicted sets are empty, $J_i = P_i = R_i = F_i = 1$. If exactly one set is empty, the undefined precision or recall ratio is assigned zero, and F1 and Jaccard similarity are also zero. F1 is averaged *after* calculating label overlap for each instance; it is neither micro-F1 nor the harmonic mean of the fold-level mean precision and recall. For the same predictions, $J_i = F_i/(2 - F_i)$ and $J_i \leq F_i$; consequently, $\overline{F}/(2 - \overline{F}) \leq \overline{J} \leq \overline{F}$. These relationships apply to predictions from a single configuration, not to table entries evaluated using different metric-specific settings.

For a nonempty reference label set, coverage is $\max_{\ell \in Y_i} |\{m : s_{im} \geq s_{i\ell}\}| - 1$. Instance AP is

$$\mathrm{AP}_i = \frac{1}{|Y_i|} \sum_{\ell \in Y_i} \frac{|\{m \in Y_i : s_{im} \geq s_{i\ell}\}|}{|\{m : s_{im} \geq s_{i\ell}\}|}. \tag{9}$$

Thus, AP averages precision over the reference labels for each instance, including all labels tied at each score threshold; it is not micro-AP pooled across cells. An instance with no reference labels receives a coverage of zero and an AP of one. Ranking loss is zero when there are no relevant–irrelevant label pairs. ROC AUC is undefined when a fold contains only one class of binary reference indicator; such values are not replaced by zero.

4.4.3. Per-Label Evaluation

Supplementary Table S2 reports per-label precision, recall, and F1 for ML-localkNN with Canberra and Euclidean distances on each activity channel. For each distance–channel pair, all three metrics use the single retained F1 configuration documented in Supplementary Table S1, without retuning. We pool the ten archived sets of held-out predictions, with each of the 241 labeled cells contributing once, and calculate per-label scores from the pooled confusion counts. These scores differ from the instance-based, unweighted fold means in the main tables. The number of positive reference instances is reported for all six labels; specificity for *forest* additionally measures recognition of the 21 cells without that reference label. Table S2 also reports the exact configuration settings and aggregation definitions.

4.4.4. Performance Comparison

Boldface identifies the best displayed result for each metric across all eleven distances and classifier–channel combinations, considering both pages of each *k*-NN table together. Ties at the displayed precision are retained, and baselines are compared separately. A *t*-test is conducted with an alpha level of 0.05 to determine if there are significant differences between the best result and the remaining results. If there are no significant differences, the results are marked in italics.

5. Results

5.1. Performance Tables

Each row identifies a distance–channel–classifier combination. The metrics are presented in two groups: Jaccard accuracy, Hamming loss, coverage, ranking loss, and average precision; and F1, recall, precision, and ROC AUC. Each *k*-NN table spans two pages, presenting the five base distances followed by the six CID-based distances, including the elastic composite. The distance abbreviations are Eucl, Manh, Log, HasD, CanD, CID-Eucl, CID-Manh, CID-Log, CID-HasD, CID-CanD, and CID-DTW.

Tables 1 and 2 report the performance of the fixed non-*k*-NN reference configurations. Arrows indicate whether higher (↑) or lower (↓) values represent better performance. On *callin*, LP has the highest mean Jaccard accuracy and F1 among these baselines, at 0.658 and 0.741, respectively, whereas RAkEL-d has the lowest Hamming loss and ranking loss, at 0.222 and 0.145. Performance among the reference configurations therefore also depends on the evaluation objective. Their fixed hyperparameters and role in contextualizing the distance comparison are described in Section 4.3; these results do not establish a ranking of classifier families.

Table 1. Baseline performance on the Trento benchmark: Jaccard accuracy, Hamming loss, coverage, ranking loss, and average precision. Highlighting follows Section 4.4.4.

Dataset	Classifier	Metric				
		Acc. (↑)	HL (↓)	Cov. (↓)	RankL (↓)	AvgP (↑)
callin	CC	0.585 ± 0.059	0.289 ± 0.043	3.618 ± 0.265	0.275 ± 0.051	0.813 ± 0.037
	LP	**0.658 ± 0.053**	*0.237 ± 0.040*	**2.762 ± 0.235**	*0.151 ± 0.033*	*0.884 ± 0.019*
	RAkEL-d	0.610 ± 0.061	**0.222 ± 0.041**	*2.766 ± 0.233*	**0.145 ± 0.031**	**0.890 ± 0.018**
	ML-CRF	0.557 ± 0.067	0.313 ± 0.048	3.050 ± 0.129	0.207 ± 0.016	0.840 ± 0.018
callout	CC	0.576 ± 0.045	0.296 ± 0.033	3.656 ± 0.206	0.285 ± 0.037	0.806 ± 0.029
	LP	*0.650 ± 0.075*	0.243 ± 0.054	*2.773 ± 0.264*	*0.154 ± 0.029*	*0.883 ± 0.020*
	RAkEL-d	0.602 ± 0.048	*0.225 ± 0.033*	*2.833 ± 0.263*	*0.157 ± 0.033*	*0.881 ± 0.024*
	ML-CRF	0.555 ± 0.042	0.314 ± 0.029	3.048 ± 0.117	0.207 ± 0.019	0.840 ± 0.018
internet	CC	0.580 ± 0.047	0.293 ± 0.031	3.657 ± 0.183	0.281 ± 0.033	0.809 ± 0.028
	LP	*0.633 ± 0.061*	0.255 ± 0.045	*2.804 ± 0.240*	*0.159 ± 0.026*	0.880 ± 0.014
	RAkEL-d	0.585 ± 0.051	*0.234 ± 0.032*	*2.838 ± 0.184*	0.161 ± 0.023	0.875 ± 0.015
	ML-CRF	0.558 ± 0.054	0.312 ± 0.036	3.041 ± 0.069	0.206 ± 0.027	0.841 ± 0.031

Table 2. Baseline performance on the Trento benchmark: F1, recall, precision, and ROC AUC.

Dataset	Classifier	Metric			
		F1 (↑)	Rec. (↑)	Prec. (↑)	ROC (↑)
callin	CC	0.683 ± 0.044	0.712 ± 0.040	0.669 ± 0.048	0.736 ± 0.047
	LP	**0.741 ± 0.040**	**0.754 ± 0.037**	0.740 ± 0.045	*0.859 ± 0.037*
	RAkEL-d	*0.728 ± 0.044*	0.683 ± 0.048	*0.812 ± 0.040*	**0.865 ± 0.033**
	ML-CRF	0.658 ± 0.053	0.685 ± 0.048	0.645 ± 0.058	0.786 ± 0.017
callout	CC	0.674 ± 0.036	0.701 ± 0.035	0.662 ± 0.037	0.729 ± 0.033
	LP	*0.734 ± 0.060*	*0.748 ± 0.065*	0.733 ± 0.054	*0.854 ± 0.035*
	RAkEL-d	*0.723 ± 0.037*	0.673 ± 0.042	**0.813 ± 0.037**	*0.854 ± 0.038*
	ML-CRF	0.656 ± 0.033	0.683 ± 0.031	0.643 ± 0.035	0.789 ± 0.018
internet	CC	0.678 ± 0.038	0.704 ± 0.032	0.666 ± 0.044	0.727 ± 0.035
	LP	0.720 ± 0.047	0.733 ± 0.050	0.720 ± 0.044	0.847 ± 0.035
	RAkEL-d	0.710 ± 0.038	0.654 ± 0.050	*0.809 ± 0.037*	0.846 ± 0.029
	ML-CRF	0.659 ± 0.044	0.687 ± 0.037	0.646 ± 0.051	0.789 ± 0.025

5.2. Distance, Classifier, and Channel Effects

Tables 3 and 4 show that the relative performance of distance functions depends on the classifier, activity channel, and evaluation objective. All comparisons below refer to the retained metric-specific configurations. For ML-localkNN on *callin*, Canberra yields a ranking loss of 0.125 ± 0.034 and an average precision of 0.905 ± 0.022, compared with 0.137 ± 0.039 and 0.900 ± 0.022, respectively, for Euclidean distance. Mean precision, however, is higher with Euclidean distance: 0.828 versus 0.821 with Canberra. Distance rankings vary across channels. In the case of ML-kNN, for *callin*, Canberra distance has a higher mean F1 score than Euclidean distance: 0.770 ± 0.043 and 0.757 ± 0.037, respectively.

For *internet*, the ordering is reversed: Euclidean distance provides 0.756 ± 0.046 while Canberra provides 0.739 ± 0.035. Within a given activity channel, the relative performance of distance functions also depends on the classifier. For *callin*, Canberra distance has the highest mean F1 score among the eleven distances for BR-kNN, ML-kNN, and LAML-kNN. In the case of ML-localkNN, CID-Euclidean has the highest mean F1 score of 0.765, while Canberra provides 0.761.

Table 3. *k*-NN performance across all eleven distances: Jaccard accuracy, Hamming loss, coverage, ranking loss, and average precision. Base distances are followed by their CID counterparts and CID-DTW10. Highlighting follows Section 4.4.4 across both pages.

Dist.	Dataset	Classifier	Metric				
			Acc. (↑)	HL (↓)	Cov. (↓)	RankL (↓)	AvgP (↑)
Eucl	*callin*	BR-kNN	*0.678 ± 0.047*	*0.212 ± 0.033*	2.704 ± 0.308	0.157 ± 0.042	0.881 ± 0.023
		ML-kNN	*0.663 ± 0.053*	*0.210 ± 0.033*	2.598 ± 0.230	*0.135 ± 0.031*	*0.901 ± 0.018*
		LAML-kNN	0.643 ± 0.059	*0.223 ± 0.039*	2.651 ± 0.234	0.165 ± 0.047	0.872 ± 0.016
		ML-localkNN	*0.662 ± 0.061*	*0.211 ± 0.043*	2.610 ± 0.263	*0.137 ± 0.039*	*0.900 ± 0.022*
	callout	BR-kNN	*0.656 ± 0.066*	0.226 ± 0.047	2.720 ± 0.195	0.159 ± 0.023	0.875 ± 0.019
		ML-kNN	0.643 ± 0.083	*0.223 ± 0.054*	2.643 ± 0.297	0.146 ± 0.037	0.888 ± 0.030
		LAML-kNN	0.636 ± 0.083	0.228 ± 0.054	2.635 ± 0.302	0.168 ± 0.045	0.873 ± 0.030
		ML-localkNN	0.640 ± 0.069	*0.223 ± 0.034*	2.638 ± 0.288	0.150 ± 0.033	0.888 ± 0.024
	internet	BR-kNN	*0.666 ± 0.059*	0.225 ± 0.033	2.862 ± 0.223	0.170 ± 0.024	0.869 ± 0.017
		ML-kNN	*0.666 ± 0.059*	0.226 ± 0.040	2.687 ± 0.191	0.147 ± 0.026	0.891 ± 0.024
		LAML-kNN	*0.666 ± 0.059*	0.226 ± 0.040	2.687 ± 0.191	0.147 ± 0.026	0.891 ± 0.024
		ML-localkNN	0.633 ± 0.053	0.227 ± 0.032	2.703 ± 0.205	0.152 ± 0.026	0.885 ± 0.024
Manh	*callin*	BR-kNN	*0.664 ± 0.055*	*0.216 ± 0.040*	2.700 ± 0.283	0.156 ± 0.042	0.881 ± 0.028
		ML-kNN	*0.665 ± 0.045*	*0.210 ± 0.034*	2.658 ± 0.216	*0.139 ± 0.043*	*0.896 ± 0.020*
		LAML-kNN	*0.652 ± 0.062*	*0.218 ± 0.045*	2.688 ± 0.244	0.169 ± 0.026	0.875 ± 0.019
		ML-localkNN	*0.663 ± 0.062*	*0.213 ± 0.027*	2.605 ± 0.276	*0.140 ± 0.037*	*0.898 ± 0.030*
	callout	BR-kNN	0.646 ± 0.077	0.231 ± 0.031	2.731 ± 0.166	0.161 ± 0.025	0.873 ± 0.021
		ML-kNN	0.640 ± 0.073	0.233 ± 0.029	2.679 ± 0.232	0.148 ± 0.030	0.886 ± 0.022
		LAML-kNN	0.634 ± 0.071	0.231 ± 0.043	2.702 ± 0.193	0.167 ± 0.034	0.872 ± 0.029
		ML-localkNN	0.631 ± 0.053	0.228 ± 0.043	2.657 ± 0.231	0.147 ± 0.036	0.884 ± 0.027
	internet	BR-kNN	*0.652 ± 0.051*	0.235 ± 0.050	2.852 ± 0.240	0.168 ± 0.028	0.873 ± 0.021
		ML-kNN	*0.652 ± 0.051*	0.236 ± 0.041	2.717 ± 0.244	0.153 ± 0.031	0.887 ± 0.025
		LAML-kNN	*0.652 ± 0.051*	0.237 ± 0.036	2.729 ± 0.150	0.155 ± 0.028	0.885 ± 0.026
		ML-localkNN	0.621 ± 0.059	0.243 ± 0.040	2.734 ± 0.237	0.154 ± 0.033	0.886 ± 0.028
Log	*callin*	BR-kNN	*0.662 ± 0.055*	*0.216 ± 0.040*	2.700 ± 0.283	0.155 ± 0.042	0.881 ± 0.029
		ML-kNN	*0.666 ± 0.044*	*0.210 ± 0.033*	2.654 ± 0.213	*0.139 ± 0.043*	*0.896 ± 0.019*
		LAML-kNN	*0.654 ± 0.064*	*0.215 ± 0.036*	2.682 ± 0.219	0.168 ± 0.039	0.875 ± 0.019
		ML-localkNN	*0.663 ± 0.052*	*0.216 ± 0.027*	2.621 ± 0.274	*0.138 ± 0.030*	*0.898 ± 0.027*
	callout	BR-kNN	0.646 ± 0.077	0.231 ± 0.031	2.722 ± 0.165	0.161 ± 0.021	0.873 ± 0.020
		ML-kNN	0.640 ± 0.073	0.232 ± 0.042	2.679 ± 0.232	0.148 ± 0.030	0.886 ± 0.022
		LAML-kNN	0.635 ± 0.071	0.231 ± 0.043	2.702 ± 0.191	0.167 ± 0.034	0.872 ± 0.029
		ML-localkNN	0.636 ± 0.069	0.229 ± 0.041	2.649 ± 0.210	0.149 ± 0.034	0.884 ± 0.032
	internet	BR-kNN	*0.652 ± 0.051*	0.235 ± 0.050	2.852 ± 0.240	0.168 ± 0.028	0.873 ± 0.021
		ML-kNN	*0.652 ± 0.051*	0.236 ± 0.041	2.717 ± 0.244	0.153 ± 0.031	0.887 ± 0.025
		LAML-kNN	*0.652 ± 0.051*	0.237 ± 0.036	2.729 ± 0.150	0.155 ± 0.028	0.885 ± 0.026
		ML-localkNN	0.620 ± 0.052	0.241 ± 0.040	2.718 ± 0.247	0.152 ± 0.031	0.886 ± 0.030
HasD	*callin*	BR-kNN	*0.670 ± 0.055*	*0.215 ± 0.039*	2.697 ± 0.290	0.155 ± 0.042	0.881 ± 0.029
		ML-kNN	*0.663 ± 0.046*	*0.213 ± 0.033*	2.649 ± 0.242	*0.140 ± 0.045*	*0.899 ± 0.020*
		LAML-kNN	*0.653 ± 0.055*	*0.218 ± 0.036*	2.690 ± 0.215	0.163 ± 0.034	0.875 ± 0.019
		ML-localkNN	*0.672 ± 0.051*	*0.209 ± 0.043*	2.621 ± 0.270	*0.139 ± 0.030*	*0.897 ± 0.021*
	callout	BR-kNN	0.646 ± 0.077	0.231 ± 0.054	2.726 ± 0.159	0.161 ± 0.025	0.873 ± 0.020
		ML-kNN	0.636 ± 0.073	0.231 ± 0.029	2.669 ± 0.178	0.147 ± 0.031	0.888 ± 0.023
		LAML-kNN	0.632 ± 0.077	0.235 ± 0.020	2.694 ± 0.188	0.169 ± 0.040	0.872 ± 0.029
		ML-localkNN	0.635 ± 0.057	0.231 ± 0.045	2.636 ± 0.209	0.146 ± 0.038	0.886 ± 0.023
	internet	BR-kNN	*0.652 ± 0.051*	0.236 ± 0.040	2.852 ± 0.240	0.168 ± 0.028	0.873 ± 0.021
		ML-kNN	*0.652 ± 0.051*	0.234 ± 0.041	2.717 ± 0.244	0.154 ± 0.031	0.885 ± 0.026
		LAML-kNN	*0.652 ± 0.051*	0.237 ± 0.036	2.729 ± 0.150	0.155 ± 0.028	0.885 ± 0.026
		ML-localkNN	0.626 ± 0.055	0.237 ± 0.038	2.707 ± 0.232	0.154 ± 0.028	0.886 ± 0.029

Table 3. *Cont.*

Dist.	Dataset	Classifier	Metric				
			Acc. (↑)	HL (↓)	Cov. (↓)	RankL (↓)	AvgP (↑)
CanD	*callin*	BR-kNN	**0.683 ± 0.064**	**0.203 ± 0.042**	2.603 ± 0.227	0.146 ± 0.036	0.891 ± 0.025
		ML-kNN	*0.681 ± 0.046*	*0.205 ± 0.035*	*2.546 ± 0.200*	*0.126 ± 0.036*	**0.905 ± 0.024**
		LAML-kNN	*0.677 ± 0.044*	*0.206 ± 0.030*	*2.568 ± 0.217*	0.154 ± 0.043	0.888 ± 0.021
		ML-localkNN	*0.671 ± 0.046*	*0.208 ± 0.034*	**2.494 ± 0.202**	**0.125 ± 0.034**	**0.905 ± 0.022**
	callout	BR-kNN	*0.665 ± 0.060*	*0.215 ± 0.041*	2.652 ± 0.216	0.148 ± 0.024	0.891 ± 0.028
		ML-kNN	*0.656 ± 0.072*	*0.217 ± 0.048*	2.637 ± 0.271	*0.140 ± 0.025*	*0.897 ± 0.026*
		LAML-kNN	*0.655 ± 0.083*	0.232 ± 0.038	2.671 ± 0.209	0.150 ± 0.044	0.891 ± 0.029
		ML-localkNN	*0.658 ± 0.079*	*0.220 ± 0.047*	2.657 ± 0.234	*0.135 ± 0.036*	*0.896 ± 0.021*
	internet	BR-kNN	0.644 ± 0.047	0.239 ± 0.040	2.880 ± 0.189	0.175 ± 0.032	0.867 ± 0.015
		ML-kNN	0.644 ± 0.047	0.245 ± 0.032	2.732 ± 0.254	0.160 ± 0.017	0.883 ± 0.018
		LAML-kNN	0.644 ± 0.047	0.245 ± 0.032	2.740 ± 0.155	0.160 ± 0.017	0.883 ± 0.018
		ML-localkNN	0.614 ± 0.051	0.240 ± 0.040	2.769 ± 0.172	0.163 ± 0.034	0.876 ± 0.020
CID-Eucl	*callin*	BR-kNN	*0.675 ± 0.048*	*0.210 ± 0.032*	2.701 ± 0.253	0.152 ± 0.042	0.884 ± 0.026
		ML-kNN	*0.669 ± 0.042*	*0.213 ± 0.046*	2.619 ± 0.241	*0.136 ± 0.032*	*0.897 ± 0.020*
		LAML-kNN	*0.659 ± 0.058*	*0.213 ± 0.038*	2.641 ± 0.231	0.161 ± 0.052	0.875 ± 0.018
		ML-localkNN	*0.670 ± 0.042*	*0.209 ± 0.044*	*2.576 ± 0.264*	*0.136 ± 0.035*	*0.899 ± 0.019*
	callout	BR-kNN	*0.666 ± 0.074*	*0.216 ± 0.053*	2.724 ± 0.220	0.160 ± 0.025	0.877 ± 0.020
		ML-kNN	*0.652 ± 0.087*	*0.221 ± 0.050*	2.642 ± 0.270	*0.142 ± 0.032*	0.891 ± 0.022
		LAML-kNN	*0.652 ± 0.080*	0.228 ± 0.056	2.658 ± 0.270	0.162 ± 0.050	0.873 ± 0.032
		ML-localkNN	0.643 ± 0.057	*0.220 ± 0.041*	2.607 ± 0.256	*0.141 ± 0.030*	0.892 ± 0.025
	internet	BR-kNN	*0.652 ± 0.051*	0.232 ± 0.039	2.835 ± 0.236	0.168 ± 0.025	0.871 ± 0.015
		ML-kNN	*0.652 ± 0.051*	0.237 ± 0.037	2.691 ± 0.255	0.152 ± 0.027	0.886 ± 0.018
		LAML-kNN	*0.652 ± 0.051*	0.237 ± 0.037	2.691 ± 0.255	0.152 ± 0.027	0.886 ± 0.018
		ML-localkNN	0.618 ± 0.059	0.238 ± 0.046	2.746 ± 0.230	0.161 ± 0.033	0.879 ± 0.021
CID-Manh	*callin*	BR-kNN	*0.661 ± 0.047*	*0.214 ± 0.041*	2.688 ± 0.283	0.151 ± 0.042	0.886 ± 0.026
		ML-kNN	*0.658 ± 0.069*	*0.216 ± 0.035*	2.629 ± 0.231	*0.138 ± 0.030*	*0.896 ± 0.022*
		LAML-kNN	*0.667 ± 0.060*	*0.212 ± 0.047*	2.652 ± 0.298	0.160 ± 0.043	0.875 ± 0.017
		ML-localkNN	*0.663 ± 0.069*	*0.214 ± 0.046*	*2.585 ± 0.285*	*0.131 ± 0.034*	*0.902 ± 0.019*
	callout	BR-kNN	*0.655 ± 0.070*	*0.221 ± 0.047*	2.712 ± 0.203	0.156 ± 0.024	0.876 ± 0.018
		ML-kNN	*0.651 ± 0.066*	*0.223 ± 0.051*	2.665 ± 0.238	0.145 ± 0.032	0.890 ± 0.025
		LAML-kNN	*0.651 ± 0.076*	0.232 ± 0.018	2.674 ± 0.246	0.165 ± 0.057	0.872 ± 0.033
		ML-localkNN	*0.654 ± 0.065*	*0.222 ± 0.040*	2.639 ± 0.242	*0.140 ± 0.025*	0.891 ± 0.018
	internet	BR-kNN	*0.653 ± 0.046*	0.227 ± 0.035	2.844 ± 0.181	0.169 ± 0.021	0.872 ± 0.017
		ML-kNN	*0.653 ± 0.046*	0.235 ± 0.032	2.689 ± 0.229	0.151 ± 0.025	0.884 ± 0.019
		LAML-kNN	*0.653 ± 0.046*	0.235 ± 0.032	2.689 ± 0.229	0.151 ± 0.025	0.884 ± 0.019
		ML-localkNN	0.622 ± 0.071	0.237 ± 0.048	2.744 ± 0.232	0.154 ± 0.028	0.880 ± 0.021
CID-Log	*callin*	BR-kNN	*0.661 ± 0.047*	*0.214 ± 0.041*	2.689 ± 0.249	0.151 ± 0.042	0.886 ± 0.026
		ML-kNN	*0.656 ± 0.070*	*0.214 ± 0.047*	2.637 ± 0.227	*0.139 ± 0.031*	*0.896 ± 0.022*
		LAML-kNN	*0.662 ± 0.064*	*0.211 ± 0.045*	2.660 ± 0.298	0.160 ± 0.042	0.875 ± 0.017
		ML-localkNN	*0.665 ± 0.066*	*0.209 ± 0.051*	*2.593 ± 0.270*	*0.132 ± 0.032*	*0.903 ± 0.019*
	callout	BR-kNN	*0.655 ± 0.070*	*0.221 ± 0.047*	2.712 ± 0.203	0.156 ± 0.024	0.876 ± 0.018
		ML-kNN	*0.651 ± 0.066*	*0.223 ± 0.050*	2.673 ± 0.215	0.146 ± 0.033	0.889 ± 0.026
		LAML-kNN	*0.651 ± 0.076*	0.232 ± 0.053	2.674 ± 0.246	0.165 ± 0.057	0.872 ± 0.033
		ML-localkNN	0.650 ± 0.062	0.224 ± 0.055	2.642 ± 0.257	*0.139 ± 0.026*	0.892 ± 0.020
	internet	BR-kNN	*0.653 ± 0.046*	0.227 ± 0.034	2.844 ± 0.181	0.169 ± 0.021	0.872 ± 0.017
		ML-kNN	*0.653 ± 0.046*	0.235 ± 0.032	2.689 ± 0.229	0.151 ± 0.025	0.884 ± 0.019
		LAML-kNN	*0.653 ± 0.046*	0.235 ± 0.032	2.689 ± 0.229	0.151 ± 0.025	0.884 ± 0.019
		ML-localkNN	0.621 ± 0.071	0.238 ± 0.047	2.750 ± 0.240	0.156 ± 0.029	0.879 ± 0.016
CID-HasD	*callin*	BR-kNN	*0.658 ± 0.046*	*0.216 ± 0.038*	2.689 ± 0.249	0.151 ± 0.042	0.886 ± 0.026
		ML-kNN	*0.658 ± 0.064*	*0.215 ± 0.034*	2.629 ± 0.217	*0.138 ± 0.029*	*0.896 ± 0.019*
		LAML-kNN	*0.663 ± 0.049*	*0.214 ± 0.039*	2.655 ± 0.309	0.161 ± 0.042	0.875 ± 0.017
		ML-localkNN	*0.660 ± 0.068*	*0.213 ± 0.049*	2.602 ± 0.275	*0.131 ± 0.038*	*0.900 ± 0.024*
	callout	BR-kNN	*0.656 ± 0.069*	*0.221 ± 0.047*	2.708 ± 0.197	0.156 ± 0.024	0.877 ± 0.019
		ML-kNN	*0.655 ± 0.064*	*0.222 ± 0.040*	2.659 ± 0.226	0.144 ± 0.028	0.890 ± 0.020
		LAML-kNN	*0.651 ± 0.076*	0.229 ± 0.038	2.669 ± 0.243	0.164 ± 0.023	0.872 ± 0.033
		ML-localkNN	*0.651 ± 0.059*	*0.220 ± 0.041*	2.649 ± 0.238	*0.141 ± 0.036*	0.890 ± 0.023
	internet	BR-kNN	0.650 ± 0.046	0.228 ± 0.034	2.839 ± 0.183	0.169 ± 0.022	0.872 ± 0.018
		ML-kNN	0.650 ± 0.046	0.238 ± 0.033	2.702 ± 0.228	0.154 ± 0.022	0.883 ± 0.018
		LAML-kNN	0.650 ± 0.046	0.238 ± 0.033	2.702 ± 0.228	0.154 ± 0.022	0.883 ± 0.018
		ML-localkNN	0.616 ± 0.073	0.243 ± 0.050	2.754 ± 0.232	0.159 ± 0.026	0.879 ± 0.014

Table 3. *Cont.*

Dist.	Dataset	Classifier	Metric				
			Acc. (↑)	HL (↓)	Cov. (↓)	RankL (↓)	AvgP (↑)
CID-CanD	*callin*	BR-kNN	*0.675 ± 0.051*	*0.207 ± 0.039*	2.615 ± 0.232	0.145 ± 0.039	0.892 ± 0.023
		ML-kNN	*0.668 ± 0.055*	*0.204 ± 0.037*	*2.564 ± 0.217*	*0.127 ± 0.030*	*0.903 ± 0.021*
		LAML-kNN	0.637 ± 0.037	*0.222 ± 0.067*	2.600 ± 0.230	0.152 ± 0.019	0.886 ± 0.014
		ML-localkNN	*0.668 ± 0.044*	*0.207 ± 0.029*	*2.557 ± 0.179*	*0.126 ± 0.031*	*0.904 ± 0.024*
	callout	BR-kNN	*0.669 ± 0.055*	*0.210 ± 0.042*	2.671 ± 0.233	0.149 ± 0.031	0.885 ± 0.018
		ML-kNN	*0.665 ± 0.047*	*0.213 ± 0.034*	*2.592 ± 0.193*	*0.133 ± 0.031*	*0.899 ± 0.025*
		LAML-kNN	*0.670 ± 0.059*	*0.209 ± 0.039*	2.629 ± 0.223	0.147 ± 0.041	0.892 ± 0.027
		ML-localkNN	*0.679 ± 0.062*	*0.204 ± 0.030*	2.628 ± 0.215	*0.134 ± 0.031*	*0.901 ± 0.019*
	internet	BR-kNN	0.638 ± 0.048	0.233 ± 0.040	2.881 ± 0.147	0.171 ± 0.020	0.870 ± 0.013
		ML-kNN	0.638 ± 0.048	0.239 ± 0.033	2.724 ± 0.239	0.157 ± 0.029	0.880 ± 0.019
		LAML-kNN	0.638 ± 0.048	0.242 ± 0.022	2.724 ± 0.239	0.158 ± 0.024	0.880 ± 0.019
		ML-localkNN	0.609 ± 0.061	0.242 ± 0.044	2.731 ± 0.160	0.155 ± 0.022	0.881 ± 0.021
CID-DTW	*callin*	BR-kNN	*0.653 ± 0.042*	*0.223 ± 0.041*	2.723 ± 0.291	0.156 ± 0.041	0.883 ± 0.025
		ML-kNN	0.649 ± 0.037	0.225 ± 0.034	2.681 ± 0.244	*0.141 ± 0.033*	0.889 ± 0.019
		LAML-kNN	0.648 ± 0.036	0.225 ± 0.025	2.682 ± 0.290	0.168 ± 0.039	0.872 ± 0.026
		ML-localkNN	*0.653 ± 0.071*	*0.220 ± 0.045*	2.641 ± 0.273	*0.139 ± 0.029*	0.893 ± 0.020
	callout	BR-kNN	*0.653 ± 0.057*	*0.224 ± 0.041*	2.704 ± 0.212	0.157 ± 0.021	0.876 ± 0.027
		ML-kNN	0.649 ± 0.067	*0.223 ± 0.048*	2.598 ± 0.257	*0.139 ± 0.041*	0.891 ± 0.030
		LAML-kNN	0.645 ± 0.063	*0.223 ± 0.030*	2.599 ± 0.239	0.161 ± 0.053	0.876 ± 0.028
		ML-localkNN	*0.653 ± 0.057*	0.227 ± 0.040	*2.582 ± 0.271*	*0.139 ± 0.040*	0.892 ± 0.024
	internet	BR-kNN	0.634 ± 0.037	0.229 ± 0.037	2.840 ± 0.235	0.163 ± 0.027	0.871 ± 0.015
		ML-kNN	0.634 ± 0.037	0.242 ± 0.033	2.742 ± 0.192	0.154 ± 0.017	0.882 ± 0.024
		LAML-kNN	0.634 ± 0.037	0.241 ± 0.034	2.757 ± 0.230	0.161 ± 0.021	0.880 ± 0.022
		ML-localkNN	0.629 ± 0.059	0.227 ± 0.039	2.767 ± 0.261	0.154 ± 0.026	0.884 ± 0.024

Table 4. *k*-NN performance across all eleven distances: F1, recall, precision, and ROC AUC. Base distances are followed by their CID counterparts and CID-DTW10. Highlighting follows Section 4.4.4 across both pages.

Dist.	Dataset	Classifier	Metric			
			F1 (↑)	Rec. (↑)	Prec. (↑)	ROC (↑)
Eucl	*callin*	BR-kNN	*0.765 ± 0.037*	0.756 ± 0.042	0.841 ± 0.062	*0.859 ± 0.031*
		ML-kNN	*0.757 ± 0.037*	0.760 ± 0.065	0.815 ± 0.034	*0.863 ± 0.034*
		LAML-kNN	*0.748 ± 0.041*	0.750 ± 0.054	0.793 ± 0.048	0.835 ± 0.051
		ML-localkNN	*0.755 ± 0.045*	*0.779 ± 0.055*	0.828 ± 0.037	*0.858 ± 0.040*
	callout	BR-kNN	0.742 ± 0.053	0.733 ± 0.045	0.843 ± 0.057	*0.852 ± 0.033*
		ML-kNN	0.740 ± 0.063	0.745 ± 0.053	0.795 ± 0.046	0.849 ± 0.040
		LAML-kNN	0.738 ± 0.065	0.737 ± 0.084	0.780 ± 0.043	0.826 ± 0.045
		ML-localkNN	0.739 ± 0.040	0.758 ± 0.063	0.810 ± 0.042	*0.850 ± 0.047*
	internet	BR-kNN	*0.756 ± 0.046*	0.767 ± 0.047	*0.868 ± 0.042*	*0.855 ± 0.034*
		ML-kNN	*0.756 ± 0.046*	0.767 ± 0.047	0.799 ± 0.042	*0.857 ± 0.035*
		LAML-kNN	*0.756 ± 0.046*	0.767 ± 0.047	0.782 ± 0.039	0.842 ± 0.025
		ML-localkNN	0.740 ± 0.051	0.757 ± 0.059	0.794 ± 0.047	*0.859 ± 0.037*
Manh	*callin*	BR-kNN	*0.753 ± 0.047*	0.744 ± 0.057	*0.853 ± 0.040*	*0.860 ± 0.041*
		ML-kNN	*0.759 ± 0.033*	0.750 ± 0.049	0.807 ± 0.051	*0.858 ± 0.049*
		LAML-kNN	*0.749 ± 0.043*	0.751 ± 0.051	0.796 ± 0.047	0.832 ± 0.049
		ML-localkNN	*0.756 ± 0.036*	0.771 ± 0.060	0.827 ± 0.044	*0.856 ± 0.034*
	callout	BR-kNN	0.736 ± 0.038	0.729 ± 0.042	0.833 ± 0.063	*0.850 ± 0.041*
		ML-kNN	0.733 ± 0.056	0.739 ± 0.064	0.787 ± 0.049	0.849 ± 0.033
		LAML-kNN	0.732 ± 0.062	0.737 ± 0.068	0.786 ± 0.055	0.835 ± 0.045
		ML-localkNN	0.734 ± 0.047	0.749 ± 0.052	0.803 ± 0.058	0.847 ± 0.044
	internet	BR-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.846 ± 0.042	0.849 ± 0.041
		ML-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.786 ± 0.044	*0.852 ± 0.038*
		LAML-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.776 ± 0.044	0.831 ± 0.033
		ML-localkNN	0.725 ± 0.057	0.759 ± 0.046	0.804 ± 0.066	*0.852 ± 0.035*

Table 4. *Cont.*

Dist.	Dataset	Classifier	Metric			
			F1 (↑)	Rec. (↑)	Prec. (↑)	ROC (↑)
Log	*callin*	BR-kNN	*0.753 ± 0.047*	0.744 ± 0.057	*0.855 ± 0.038*	*0.860 ± 0.037*
		ML-kNN	*0.760 ± 0.032*	0.754 ± 0.043	0.806 ± 0.054	*0.858 ± 0.049*
		LAML-kNN	*0.751 ± 0.043*	0.751 ± 0.051	0.800 ± 0.042	0.833 ± 0.051
		ML-localkNN	*0.759 ± 0.033*	0.770 ± 0.053	0.822 ± 0.048	*0.858 ± 0.032*
	callout	BR-kNN	0.736 ± 0.038	0.729 ± 0.042	0.833 ± 0.063	*0.850 ± 0.041*
		ML-kNN	0.733 ± 0.051	0.739 ± 0.064	0.789 ± 0.047	0.848 ± 0.044
		LAML-kNN	0.728 ± 0.048	0.737 ± 0.068	0.786 ± 0.055	0.834 ± 0.044
		ML-localkNN	0.735 ± 0.042	0.751 ± 0.053	0.802 ± 0.054	0.848 ± 0.046
	internet	BR-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.846 ± 0.042	0.849 ± 0.041
		ML-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.783 ± 0.044	*0.851 ± 0.038*
		LAML-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.775 ± 0.042	0.833 ± 0.024
		ML-localkNN	0.727 ± 0.038	0.766 ± 0.050	0.803 ± 0.057	*0.850 ± 0.036*
HasD	*callin*	BR-kNN	*0.760 ± 0.044*	0.749 ± 0.049	*0.855 ± 0.038*	*0.860 ± 0.046*
		ML-kNN	*0.754 ± 0.037*	0.754 ± 0.055	0.810 ± 0.047	*0.860 ± 0.043*
		LAML-kNN	*0.753 ± 0.041*	0.754 ± 0.054	0.792 ± 0.046	0.836 ± 0.042
		ML-localkNN	*0.762 ± 0.044*	*0.784 ± 0.064*	0.825 ± 0.043	*0.861 ± 0.039*
	callout	BR-kNN	0.736 ± 0.037	0.729 ± 0.042	0.829 ± 0.069	*0.851 ± 0.029*
		ML-kNN	0.730 ± 0.052	0.732 ± 0.066	0.782 ± 0.068	0.849 ± 0.030
		LAML-kNN	0.732 ± 0.062	0.737 ± 0.068	0.780 ± 0.046	0.833 ± 0.045
		ML-localkNN	0.733 ± 0.042	0.753 ± 0.063	0.800 ± 0.062	0.847 ± 0.038
	internet	BR-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.846 ± 0.042	0.847 ± 0.040
		ML-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.786 ± 0.053	0.849 ± 0.037
		LAML-kNN	*0.746 ± 0.044*	0.760 ± 0.048	0.773 ± 0.043	0.832 ± 0.033
		ML-localkNN	0.733 ± 0.047	0.762 ± 0.053	0.794 ± 0.066	*0.852 ± 0.039*
CanD	*callin*	BR-kNN	**0.772 ± 0.054**	0.764 ± 0.058	**0.873 ± 0.043**	**0.872 ± 0.045**
		ML-kNN	*0.770 ± 0.043*	0.768 ± 0.050	0.812 ± 0.035	*0.870 ± 0.031*
		LAML-kNN	*0.769 ± 0.038*	0.760 ± 0.047	0.804 ± 0.052	0.846 ± 0.046
		ML-localkNN	*0.761 ± 0.042*	**0.797 ± 0.039**	0.821 ± 0.053	*0.870 ± 0.037*
	callout	BR-kNN	*0.756 ± 0.046*	0.751 ± 0.067	0.844 ± 0.055	*0.866 ± 0.030*
		ML-kNN	*0.747 ± 0.049*	0.751 ± 0.067	0.801 ± 0.052	*0.862 ± 0.031*
		LAML-kNN	*0.747 ± 0.066*	0.751 ± 0.067	0.781 ± 0.054	0.838 ± 0.026
		ML-localkNN	*0.749 ± 0.054*	0.753 ± 0.060	0.820 ± 0.062	*0.863 ± 0.037*
	internet	BR-kNN	0.733 ± 0.049	0.753 ± 0.041	0.843 ± 0.042	0.846 ± 0.030
		ML-kNN	0.739 ± 0.035	0.753 ± 0.041	0.772 ± 0.036	0.843 ± 0.036
		LAML-kNN	0.739 ± 0.035	0.753 ± 0.041	0.769 ± 0.038	0.828 ± 0.017
		ML-localkNN	0.724 ± 0.042	0.721 ± 0.065	0.780 ± 0.067	0.843 ± 0.035
CID-Eucl	*callin*	BR-kNN	*0.764 ± 0.038*	0.754 ± 0.047	*0.863 ± 0.037*	*0.864 ± 0.043*
		ML-kNN	*0.760 ± 0.033*	0.765 ± 0.041	0.797 ± 0.043	*0.860 ± 0.038*
		LAML-kNN	*0.758 ± 0.045*	0.761 ± 0.049	0.795 ± 0.047	0.840 ± 0.056
		ML-localkNN	*0.765 ± 0.037*	*0.787 ± 0.048*	0.807 ± 0.033	*0.864 ± 0.042*
	callout	BR-kNN	*0.752 ± 0.058*	0.738 ± 0.049	*0.854 ± 0.055*	*0.855 ± 0.035*
		ML-kNN	*0.747 ± 0.048*	0.742 ± 0.074	0.793 ± 0.057	*0.854 ± 0.030*
		LAML-kNN	*0.746 ± 0.063*	0.758 ± 0.071	0.781 ± 0.024	0.837 ± 0.060
		ML-localkNN	*0.746 ± 0.055*	0.761 ± 0.067	0.819 ± 0.044	*0.855 ± 0.051*
	internet	BR-kNN	0.744 ± 0.038	0.758 ± 0.048	0.839 ± 0.048	*0.851 ± 0.030*
		ML-kNN	0.744 ± 0.038	0.758 ± 0.048	0.784 ± 0.050	0.848 ± 0.022
		LAML-kNN	0.744 ± 0.038	0.758 ± 0.048	0.776 ± 0.052	0.839 ± 0.026
		ML-localkNN	0.722 ± 0.041	0.741 ± 0.046	0.788 ± 0.041	0.846 ± 0.029
CID-Manh	*callin*	BR-kNN	*0.755 ± 0.047*	0.749 ± 0.054	*0.865 ± 0.046*	*0.866 ± 0.041*
		ML-kNN	*0.757 ± 0.051*	0.762 ± 0.059	0.800 ± 0.058	*0.864 ± 0.039*
		LAML-kNN	*0.764 ± 0.044*	0.772 ± 0.051	0.802 ± 0.056	0.844 ± 0.051
		ML-localkNN	*0.754 ± 0.063*	0.773 ± 0.055	0.819 ± 0.047	*0.869 ± 0.033*
	callout	BR-kNN	0.745 ± 0.056	0.732 ± 0.041	0.850 ± 0.048	*0.857 ± 0.033*
		ML-kNN	0.745 ± 0.052	0.746 ± 0.074	0.795 ± 0.057	*0.853 ± 0.029*
		LAML-kNN	0.744 ± 0.062	0.750 ± 0.071	0.784 ± 0.051	0.836 ± 0.032
		ML-localkNN	*0.750 ± 0.062*	0.766 ± 0.046	0.807 ± 0.068	*0.856 ± 0.038*
	internet	BR-kNN	0.745 ± 0.035	0.760 ± 0.045	0.839 ± 0.043	0.849 ± 0.029
		ML-kNN	0.745 ± 0.035	0.760 ± 0.045	0.789 ± 0.048	0.843 ± 0.027
		LAML-kNN	0.745 ± 0.035	0.760 ± 0.045	0.784 ± 0.041	0.840 ± 0.021
		ML-localkNN	0.728 ± 0.056	0.736 ± 0.040	0.775 ± 0.064	0.849 ± 0.031

Table 4. *Cont.*

Dist.	Dataset	Classifier	Metric			
			F1 (↑)	Rec. (↑)	Prec. (↑)	ROC (↑)
CID-Log	*callin*	BR-kNN	*0.755 ± 0.047*	0.749 ± 0.054	*0.865 ± 0.046*	*0.866 ± 0.041*
		ML-kNN	*0.756 ± 0.053*	0.761 ± 0.060	0.800 ± 0.054	*0.864 ± 0.039*
		LAML-kNN	*0.760 ± 0.047*	0.770 ± 0.052	0.805 ± 0.060	0.844 ± 0.051
		ML-localkNN	*0.755 ± 0.062*	0.774 ± 0.055	0.815 ± 0.051	*0.870 ± 0.032*
	callout	BR-kNN	0.745 ± 0.056	0.732 ± 0.041	0.850 ± 0.048	*0.856 ± 0.033*
		ML-kNN	0.745 ± 0.052	0.746 ± 0.074	0.794 ± 0.058	*0.853 ± 0.029*
		LAML-kNN	0.744 ± 0.062	0.750 ± 0.071	0.784 ± 0.050	0.835 ± 0.031
		ML-localkNN	*0.746 ± 0.062*	0.768 ± 0.046	0.805 ± 0.058	*0.856 ± 0.038*
	internet	BR-kNN	0.745 ± 0.035	0.760 ± 0.045	0.839 ± 0.043	0.849 ± 0.029
		ML-kNN	0.745 ± 0.035	0.760 ± 0.045	0.789 ± 0.036	0.843 ± 0.027
		LAML-kNN	0.745 ± 0.035	0.760 ± 0.045	0.784 ± 0.041	0.840 ± 0.021
		ML-localkNN	0.727 ± 0.056	0.735 ± 0.039	0.778 ± 0.063	0.850 ± 0.031
CID-HasD	*callin*	BR-kNN	*0.753 ± 0.048*	0.748 ± 0.054	*0.867 ± 0.046*	*0.866 ± 0.041*
		ML-kNN	*0.757 ± 0.050*	0.763 ± 0.060	0.801 ± 0.059	*0.863 ± 0.038*
		LAML-kNN	*0.761 ± 0.037*	0.773 ± 0.050	0.797 ± 0.062	0.844 ± 0.050
		ML-localkNN	*0.751 ± 0.051*	0.773 ± 0.056	0.814 ± 0.043	*0.868 ± 0.033*
	callout	BR-kNN	*0.745 ± 0.056*	0.732 ± 0.041	0.850 ± 0.048	*0.856 ± 0.034*
		ML-kNN	*0.749 ± 0.052*	0.751 ± 0.072	0.793 ± 0.059	*0.855 ± 0.031*
		LAML-kNN	0.736 ± 0.039	0.749 ± 0.073	0.790 ± 0.044	0.839 ± 0.029
		ML-localkNN	*0.748 ± 0.052*	0.776 ± 0.046	0.804 ± 0.058	*0.851 ± 0.038*
	internet	BR-kNN	0.742 ± 0.035	0.757 ± 0.045	0.839 ± 0.043	0.849 ± 0.029
		ML-kNN	0.742 ± 0.035	0.757 ± 0.045	0.784 ± 0.043	0.842 ± 0.020
		LAML-kNN	0.742 ± 0.035	0.757 ± 0.045	0.780 ± 0.048	0.839 ± 0.020
		ML-localkNN	0.722 ± 0.058	0.730 ± 0.047	0.777 ± 0.036	0.845 ± 0.029
CID-CanD	*callin*	BR-kNN	*0.764 ± 0.041*	0.754 ± 0.038	*0.864 ± 0.046*	**0.872 ± 0.040**
		ML-kNN	*0.762 ± 0.043*	0.758 ± 0.032	0.808 ± 0.042	*0.870 ± 0.037*
		LAML-kNN	0.740 ± 0.047	0.735 ± 0.034	0.784 ± 0.055	0.848 ± 0.049
		ML-localkNN	*0.763 ± 0.036*	0.770 ± 0.049	0.808 ± 0.060	*0.870 ± 0.036*
	callout	BR-kNN	*0.759 ± 0.046*	0.751 ± 0.053	*0.859 ± 0.046*	*0.868 ± 0.031*
		ML-kNN	*0.759 ± 0.037*	0.769 ± 0.078	0.797 ± 0.048	*0.866 ± 0.035*
		LAML-kNN	*0.762 ± 0.048*	0.763 ± 0.045	0.798 ± 0.043	0.840 ± 0.041
		ML-localkNN	*0.771 ± 0.051*	0.774 ± 0.051	0.819 ± 0.031	*0.866 ± 0.036*
	internet	BR-kNN	0.729 ± 0.047	0.744 ± 0.042	0.835 ± 0.039	0.846 ± 0.027
		ML-kNN	0.732 ± 0.047	0.744 ± 0.042	0.788 ± 0.043	0.844 ± 0.014
		LAML-kNN	0.731 ± 0.036	0.744 ± 0.042	0.783 ± 0.028	0.833 ± 0.021
		ML-localkNN	0.718 ± 0.037	0.733 ± 0.037	0.791 ± 0.056	0.844 ± 0.029
CID-DTW	*callin*	BR-kNN	0.744 ± 0.032	0.733 ± 0.059	0.848 ± 0.051	*0.863 ± 0.040*
		ML-kNN	*0.748 ± 0.031*	0.756 ± 0.037	0.801 ± 0.056	*0.856 ± 0.036*
		LAML-kNN	0.744 ± 0.038	0.753 ± 0.037	0.796 ± 0.058	0.828 ± 0.045
		ML-localkNN	*0.747 ± 0.046*	0.766 ± 0.037	0.823 ± 0.032	*0.859 ± 0.030*
	callout	BR-kNN	0.742 ± 0.050	0.738 ± 0.057	*0.852 ± 0.055*	*0.857 ± 0.039*
		ML-kNN	0.741 ± 0.040	0.748 ± 0.048	0.785 ± 0.046	*0.855 ± 0.030*
		LAML-kNN	*0.746 ± 0.033*	0.753 ± 0.040	0.793 ± 0.030	0.828 ± 0.036
		ML-localkNN	0.741 ± 0.043	0.772 ± 0.058	0.799 ± 0.064	*0.854 ± 0.049*
	internet	BR-kNN	0.727 ± 0.030	0.746 ± 0.037	0.824 ± 0.057	*0.853 ± 0.025*
		ML-kNN	0.731 ± 0.029	0.746 ± 0.037	0.786 ± 0.042	*0.851 ± 0.021*
		LAML-kNN	0.731 ± 0.029	0.746 ± 0.037	0.768 ± 0.042	0.840 ± 0.023
		ML-localkNN	0.729 ± 0.037	0.742 ± 0.060	0.784 ± 0.038	0.848 ± 0.029

The effects of complexity correction vary even when the same neighborhood size is used within each comparison. For ML-kNN on *callin*, Euclidean and CID-Euclidean distances yield mean F1 scores of 0.757 and 0.760, respectively, both with $k = 5$. On *internet*, the corresponding means are 0.756 and 0.744, both with $k = 1$. The direction of the difference therefore varies across channels for the same base–CID pair. Results for the elastic distance refer specifically to CID-DTW10, which combines temporal warping and complexity correction.

5.3. Per-Label Performance

Supplementary Table S2 shows that the relative performance of the Canberra and Euclidean distances varies across labels and channels. On *callin*, F1 for *work* rounds to 0.658 for both distances: Canberra has higher recall (0.591 versus 0.568) but lower precision (0.743 versus 0.781). Canberra has a lower F1 for *high density* (0.514 versus 0.571) but a higher F1 for *low density* (0.618 versus 0.576). F1 for *work* is also lower with Canberra on *callout* (0.571 versus 0.622) and *internet* (0.478 versus 0.560). Favorable aggregate and ranking scores therefore do not imply a consistent advantage in detecting minority labels.

The recall for *forest* varies from 0.968 to 0.991 for all six configurations, while specificity varies from 0.238 to 0.524. For *callin*, the Euclidean distance method is able to correctly recognize 9 out of 21 non-forest cells, while the Canberra distance recognizes 11, resulting in specificities of 0.429 and 0.524, respectively. The high sensitivity for the common class thus exists in combination with the limited specificity for the negative class. This analysis uses the retained F1 configurations and their pooled out-of-fold predictions.

5.4. Average Distance Ranks

Figure 3 illustrates the results of the above comparison through mean distance ranks. The eleven distances are ranked individually for each evaluation metric for each classifier–channel pair based on the mean fold scores presented in the tables, with tied values receiving mean ranks. Each subfigure presents the mean ranks of distances for one classifier over the three activity channels.

Radar charts depict favorable mean ranks for Canberra-family distances with variation across classifiers. Overall, Canberra obtains the best mean rank in terms of accuracy and F1, while CID–Canberra obtains the best mean rank in terms of ROC AUC, coverage, ranking loss, and average precision. Other distances are also competitive for certain tasks: Euclidean distance performs the best in terms of precision, and CID–Manhattan does so in terms of recall. CID-DTW10 obtains poor rankings for accuracy and F1, especially in combination with BR-kNN and ML-kNN, but its rankings differ depending on the metric and classifier.

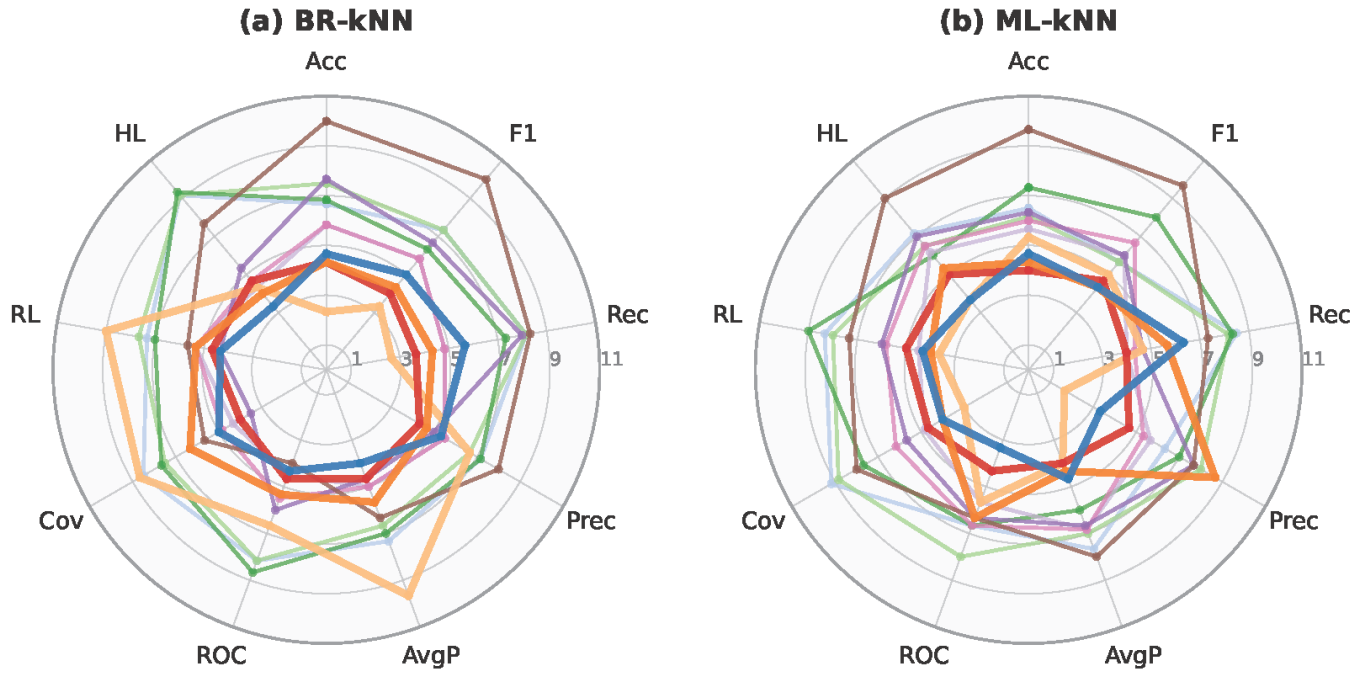


Figure 3. *Cont.*

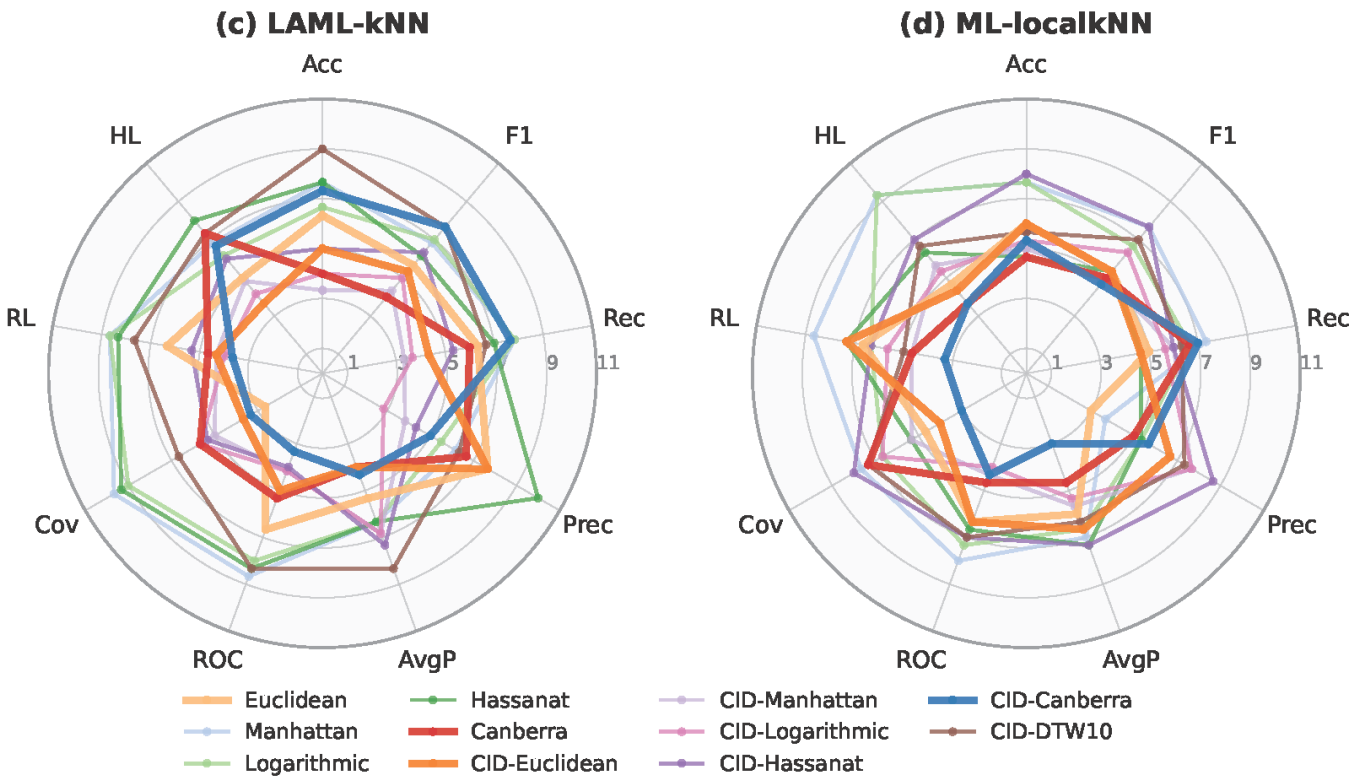


Figure 3. Radar graphs for the mean distance ranks of the nine evaluation measures in Trento: (**a**) BR-kNN; (**b**) ML-kNN; (**c**) LAML-kNN; and (**d**) ML-localkNN. Ranks are averaged across incoming calls, outgoing calls, and internet activity channels. The slightly bolder lines show the Euclidean and Canberra distances and their corresponding CID variants; the other lines are slightly more transparent. Rank 1 is the best rank.

5.5. Pairwise Profile Examples

Figure 4 provides a complementary illustration of the distance values for selected cell pairs from the *callin* dataset. The distances provide different numerical values for the same profile pairs, demonstrating the importance of proportional change, amplitude difference, and complexity scaling. Since the distance families operate on different numerical scales, the fact that one family has a larger range of values does not mean better classification of the data.

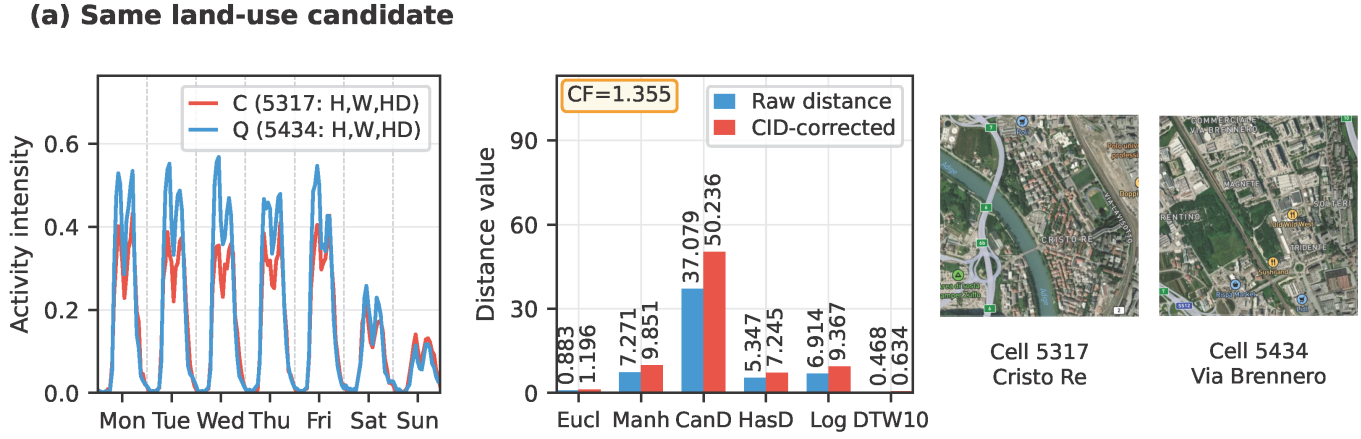


Figure 4. *Cont.*

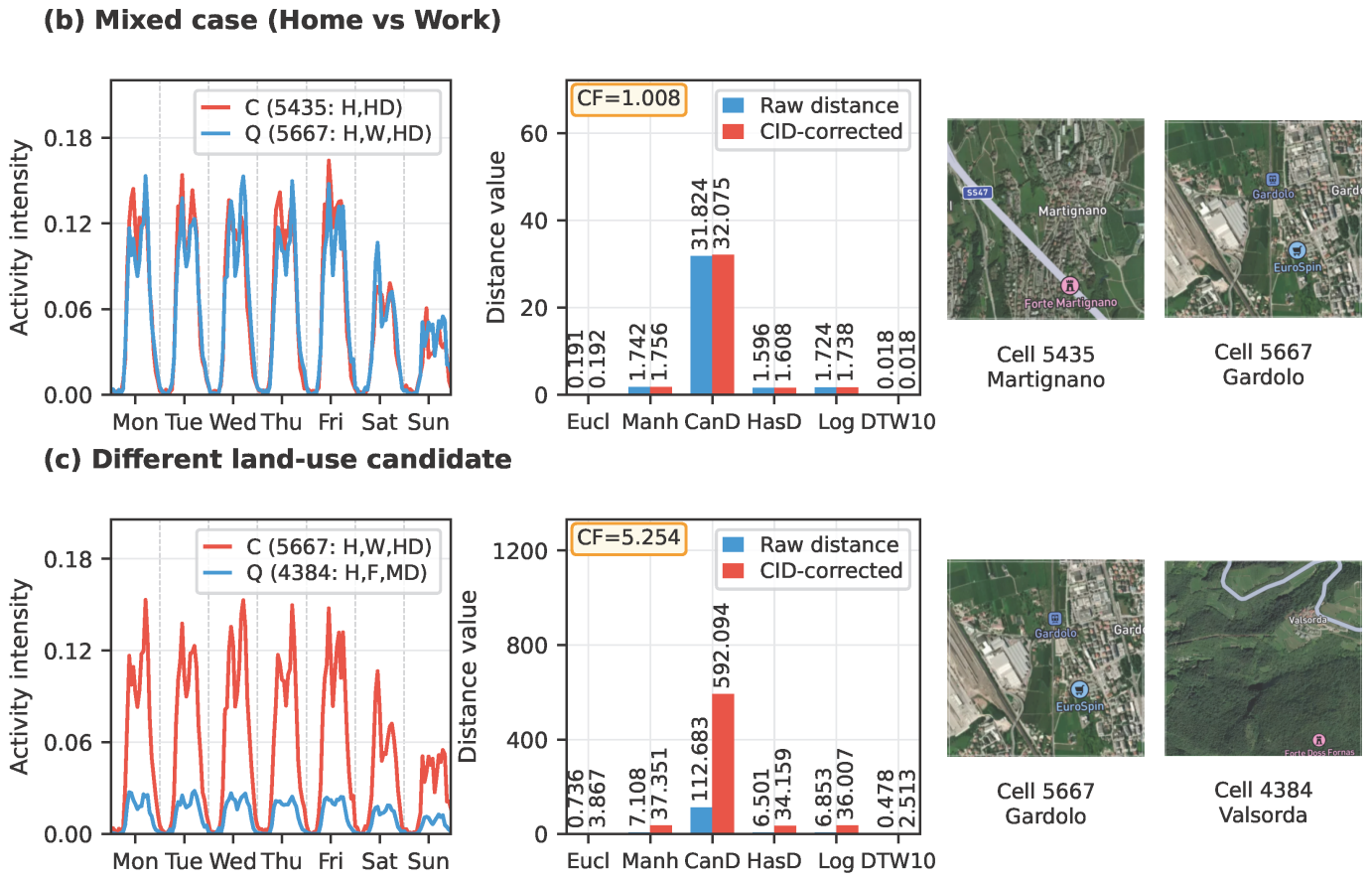


Figure 4. Pairwise distances of selected *callin* cell pairs in Trento: (**a**) a potential pair with similar land use; (**b**) residential versus mixed residential–work land use; and (**c**) a candidate pair with dissimilar land use. Each row consists of weekly activity profiles, the base and corrected CID distance values, and satellite images. CF is the CID correction factor. Profile legends identify the cells and their labels: H, home; W, work; F, forest; and HD, MD, and LD, high, medium, and low density, respectively.

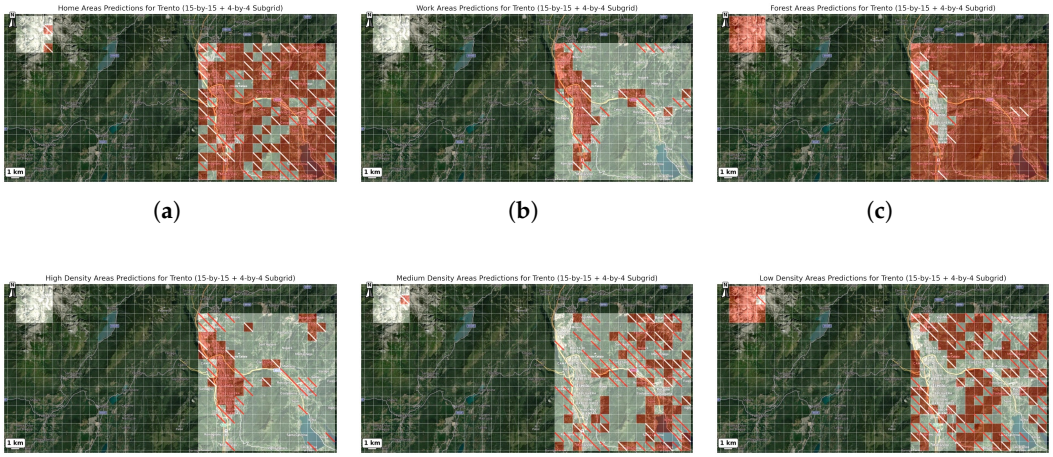
5.6. Spatial Prediction Patterns

We conclude the results section with a qualitative spatial analysis of predictions generated by ML-localkNN with Canberra distance, a combination that performs well in the benchmark. Manually labeled reference classes in Figure 2 demonstrate a spatial gradient, where the *forest* label predominates in peripheral areas, whereas high density labels are concentrated in urban areas.

Figure 5 shows predictions per label for the 15-by-15 and 4-by-4 study subgrids. Figure 6 displays the selected outputs on the full provincial grid.

The comparison between the reference labels (Figure 2) and the predictions (Figure 5) shows overall agreement for certain labels and differences in transition zones. The maps supplement the summary statistics by showing the locations and types of prediction errors. The comparison on a per-label basis in Supplementary Table S2 uses the retained F1 configurations and is analyzed separately from these qualitative map outputs. Neither analysis alone explains the observed effects of distance choice.

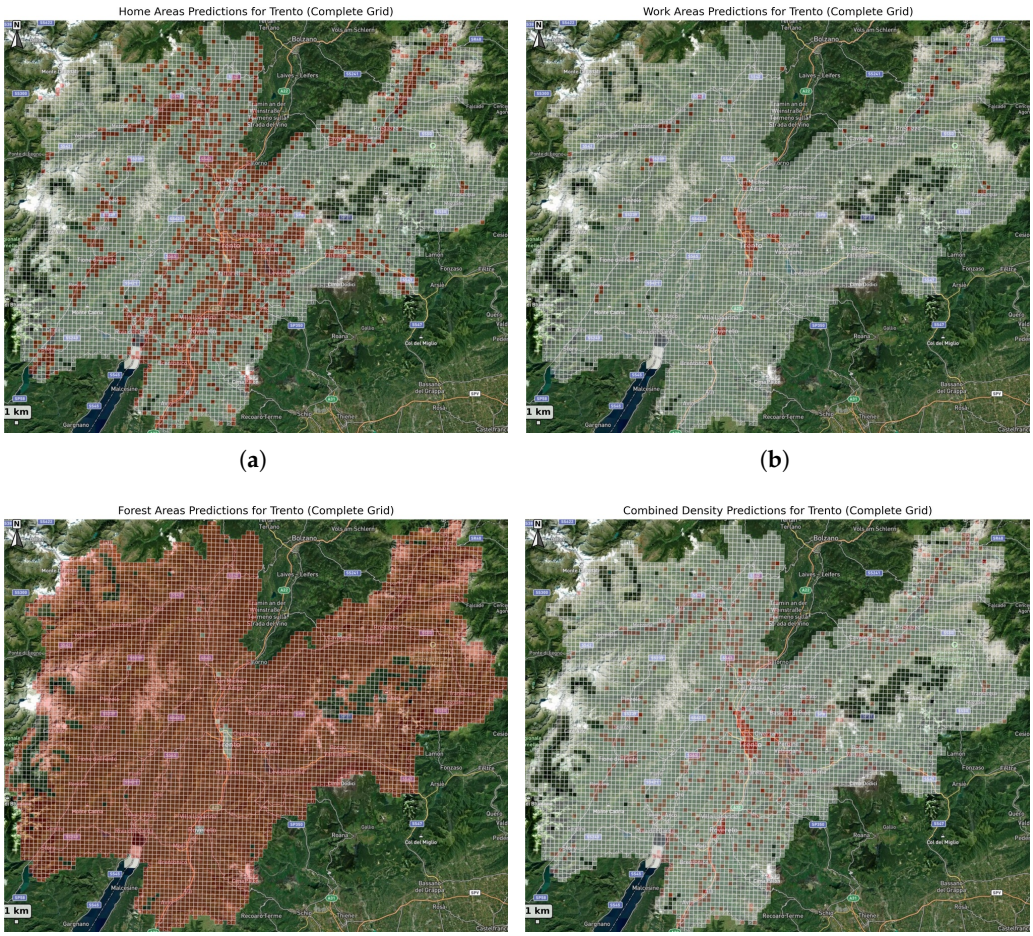
The full-grid predictions in Figure 6 demonstrate how the chosen model translates weekly activity into regional label patterns. This extends the spatial interpretation beyond the annotated study region and highlights candidate areas for collecting additional reference labels. Since the reference labels are not available for the rest of the provincial grid cells, spatial consistency alone does not validate the predictive power of the model in these regions.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

Figure 5. Spatial predictions for the Trento study area using Canberra distance with ML-localkNN: (**a**) home; (**b**) work; (**c**) forest; (**d**) high density; (**e**) medium density; and (**f**) low density. Red and white shading indicate positive and negative predictions, respectively. Diagonal lines represent prediction errors with color codes based on reference labels (red, positive; white, negative).



(**a**) (**b**)

(**c**) (**d**)

Figure 6. Predictions of the chosen Trento configuration over the entire provincial grid with Canberra distance for ML-localkNN: (**a**) home; (**b**) work; (**c**) forest; and (**d**) density. Panel (**d**) shows red, light red, and white for high, medium, and low density, respectively.

6. Discussion

The benchmark shows that the relative performance of distance functions is influenced by the choice of multilabel classifier, activity channel, and evaluation goal. This section evaluates the interaction of similarity measures with these factors in the case of Trentino and its relevance to future land use analysis.

6.1. Distance Choice Across Classifiers and Activity Channels

Canberra-family distances yield favorable ranking scores in several configurations, including an average precision of 0.905 and a ranking loss of 0.125 for ML-localkNN on incoming calls. Canberra weights each component difference relative to the magnitudes being compared, giving proportional differences greater influence than Euclidean distance does. This distinction is relevant to CDR-derived land-use signatures: Pei et al. combine hourly relative activity patterns with total call volume, treating temporal patterns and activity magnitude as complementary information [2]. Our results are consistent with a useful role for proportional differences, but they do not establish which hours or land-use labels account for the observed ranking scores. Comparing alternative normalization methods and feature contributions would help assess this interpretation.

The reversal in the relative performance of the Canberra and Euclidean distances for ML-kNN between incoming-call and internet activity (Section 5.2) also suggests that a preferred distance may not transfer directly across channels. Selection of the classifier is another important factor: Canberra outperforms the other distances on incoming calls for BR-kNN, ML-kNN, and LAML-kNN, but CID–Euclidean achieves the highest mean F1 score for ML-localkNN. ML-kNN produces the label predictions based on the frequency of labels in the chosen neighborhood [8], whereas local-k methods allow adapting the size of the neighborhood to the surrounding training instances [25]. This explains the necessity of investigating both the distance and neighborhood choice together. The differences discussed above are related to the considered configurations; their attribution to a particular internal mechanism requires investigation of the chosen neighbors and neighborhoods.

Finally, the goals of the evaluation may influence these comparisons. Favorable ranking performance for Canberra with ML-localkNN on incoming calls coexists with the high mean precision of Euclidean distance. The work of Bogatinovski et al. provides an explanation of why multilabel comparisons should take into account not one but several criteria of evaluation [15]. In land-use problems, ranking of possible labels and assignment of binary label sets are different tasks.

6.2. Complexity Correction and the Elastic Comparison

CID was proposed by Batista et al. to handle distortions related to complexity in time series distances, noting that the choice of invariances depends on the application domain [12]. The effects of CID on performance in Trentino are mixed. For ML-kNN, CID–Euclidean has a slightly better mean F1 score for incoming calls compared to Euclidean distance, while having a worse mean score for internet traffic. The same k is considered for each comparison. Thus, the rescaling of complexity can either improve or diminish the relevance of the neighborhood depending on the channel. This could be due to the fact that the variation in a weekly profile is determined not only by measurement effects but also by functionally relevant activity.

CID-DTW10's worse ranking compared to several pointwise measures also deserves attention in this particular case. According to Bagnall et al., there is a distinction between classification problems where the timing of a pattern is important and those where a pattern can move [14]. They also emphasize the significance of choosing an appropriate window in DTW-based comparisons. Our profiles have a common clock-time grid, and thus retaining

the timing of peaks could account for the performance of pointwise measures. However, CID-DTW10 combines constrained temporal warping with complexity correction, making it impossible to isolate the contribution of either component from the present results. Further experiments comparing standard DTW, CID-DTW, and different warping window sizes would be necessary to determine whether the observed performance differences arise from temporal alignment, complexity correction, or their interaction.

6.3. Regional Interpretation and Practical Use

The chosen labels reflect a combination of functional land use, built-up density, and dominant natural cover. This label set makes it possible for the benchmark to reflect a combination of settlements and forested areas. These types of labels need to be validated in a different manner. As seen in Supplementary Table S2, recall is high for *forest*, but specificity is poor. On the other hand, the minority labels *work* and *high density* have low F1 values and do not demonstrate any clear benefit of using Canberra. Although the *forest* label is present in 91.3% of the annotated cells, the evaluated configurations show limited specificity, correctly identifying only 5 to 11 of the 21 non-forest examples. This highlights the difficulty of distinguishing the minority negative class despite the high recall achieved for forest.

In practice, these maps may be useful for evaluating the mixed land use predictions, delineating transition zones for annotation, and guiding the collection of additional reference data. Spatial coherence alone does not prove the predictive validity of the model. Karasiak et al. illustrate that the dependency between training and testing locations may artificially increase the classification performance of the map, making geographically distant observations necessary for assessing the model's transferability [38]. In this study, predictions outside the annotated grid are considered exploratory, although they may provide a general overview of spatial patterns across the province.

6.4. Scope and Priorities for Validation

The current work addresses comparisons within a single region using an identical evaluation procedure. The entire dataset used in this study was scaled before the folds were formed, providing a common input scale for all configurations being tested. Yet, the estimation of such a scale using out-of-sample inputs might result in data leakage at the preprocessing stage and hence lead to bias in cross-validation estimates. The impact of such a preprocessing step on both absolute performance and distance estimates has not been quantified, and thus, the current comparisons are still conditional. When deploying the algorithm in practice, it is necessary to estimate scaling parameters using the training data only.

Iterative stratification maintains label distribution but not spatial separation; generalization to new locations therefore requires further evaluation [37,38]. Spatial validation [39], tuning, ablation analysis, and runtime analysis constitute complementary additions to the benchmark.

Due to differences in tuning of the reference classifiers and the k-NN variants tuned for each individual metric, care must be taken when interpreting comparisons between the classifier families. Likewise, CID-DTW10 combines the use of a constant warping window size with complexity correction. Comparisons with DTW and other warping window sizes may help to understand the contribution of each factor to the performance observed in this scenario.

The neighborhood configurations depend on the metrics and were chosen and evaluated in the same ten folds. Tables 3 and 4 can thus be understood as an empirical upper-bound benchmark, possibly with selection bias, rather than the performance of a

production model [48]. Because of the largely unchanged historical parameterization, the reported configurations are not necessarily optimal. A production model would have to be tuned and evaluated separately. Supplementary Table S1 contains the configuration used for each table entry.

A replication of the experiment in other regions and times would allow us to check whether the ranking of the distances is consistent when applied to different functional mixes and rhythms of activity. Label-specific or learned distances are other possibilities for further investigation, but their evaluation would still have to keep the separation between preprocessing, selection, and evaluation.

7. Conclusions

In this research, eleven distances were compared across four multilabel *k*-NN classifiers and three mobile network traffic channels in Trentino. The best-performing distance depended on the classifier, channel, and criterion used. The family of Canberra distances performed well in several scenarios, especially when ranking criteria were used on incoming call traffic. Euclidean distance also demonstrated strong performance. Per-label analysis revealed differences in the detection of work and high-density cells. The advantages of CID were also mixed, and CID-DTW10 generally performed worse than several pointwise alternatives, although these results do not isolate the effects of temporal warping from complexity correction.

The framework serves as a basis for the exploratory analysis of regions that have a combination of urban areas and forests. Prediction maps may support further model analysis and labeling. However, predictions outside of the labeled grid will need to be validated independently. For future research, some of the priorities include the estimation of preprocessing parameters in training folds, the separation of parameter selection from validation, geographical validation, and replication of the analysis in different regions and observation periods. Independent land cover reference data would improve label-based validation, while comparison with standard DTW and other window sizes could allow us to distinguish warping and complexity correction effects.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/app162111040/s1, Supplementary Table S1: configuration parameters for Tables 3 and 4. Supplementary Table S2: per-label performance of ML-localkNN.

Author Contributions: M.M.-H., J.M.C.-M., and D.O.-B. contributed to conceptualization, methodology, validation, and writing—review and editing. M.M.-H. contributed to software, formal analysis, investigation, data curation, and writing—original draft preparation. J.M.C.-M. and D.O.-B. contributed to supervision.All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Spanish Ministry of Science and Innovation under grant number PID2022-141869NB-I00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are derived from publicly available resources. The original Telecom Italia Big Data Challenge data are available from the source cited in the manuscript. The derived datasets and code used to generate the reported results are available in the following repository: https://doi.org/10.6084/m9.figshare.32200962.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Toole, J.L.; Ulm, M.; González, M.C.; Bauer, D. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*; ACM: New York, NY, USA, 2012; UrbComp '12; pp. 1–8. https://doi.org/10.1145/2346496.2346498.
2. Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.L.; Li, T.; Zhou, C. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. https://doi.org/10.1080/13658816.2014.913794.
3. Soto, V.; Frías-Martínez, E. Robust land use characterization of urban landscapes using cell phone data. In *Proceedings of the 1st Workshop on Pervasive Urban Applications*; ACM: New York, NY, USA, 2011; pp. 151–160.
4. Louail, T.; Lenormand, M.; Cantu Ros, O.G.; Picornell, M.; Herranz, R.; Frias-Martinez, E.; Ramasco, J.J.; Barthelemy, M. From mobile phone data to the spatial structure of cities. *Sci. Rep.* **2014**, *4*, 5276. https://doi.org/10.1038/srep05276.
5. Barlacchi, G.; De Nadai, M.; Larcher, R.; Casella, A.; Chitic, C.; Torrisi, G.; Antonelli, F.; Vespignani, A.; Pentland, A.; Lepri, B. A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Sci. Data* **2015**, *2*, 150055. https://doi.org/10.1038/sdata.2015.55.
6. Mendoza-Hurtado, M.; Romero-del Castillo, J.A.; Ortiz-Boyer, D. MAPLID: A new multi-label approach for place identification using data supplied by mobile network operators. *Int. J. Geogr. Inf. Sci.* **2026**, *40*, 3171–3194. https://doi.org/10.1080/13658816.2026.2617932.
7. Mendoza-Hurtado, M.; Cerruela-García, G.; Ortiz-Boyer, D. A Supervised Approach for Land Use Identification in Trento Using Mobile Phone Data as an Alternative to Unsupervised Clustering Techniques. *Appl. Sci.* **2025**, *15*, 1753. https://doi.org/10.3390/app15041753.
8. Zhang, M.L.; Zhou, Z.H. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.* **2007**, *40*, 2038–2048. https://doi.org/10.1016/j.patcog.2006.12.019.
9. Spyromitros-Xioufis, E.; Tsoumakas, G.; Vlahavas, I. An Empirical Study of Lazy Multilabel Classification Algorithms. In *Proceedings of the Artificial Intelligence: Methods and Applications (5th Hellenic Conference on Artificial Intelligence, SETN 2008), Syros, Greece, 2–4 October 2008*; Lecture Notes in Artificial Intelligence (LNAI); Springer: Berlin/Heidelberg, Germany, 2008; Volume 5138, pp. 401–406. https://doi.org/10.1007/978-3-540-87881-0_40.
10. Lance, G.N.; Williams, W.T. Computer programs for hierarchical polythetic classification ("similarity analyses"). *Comput. J.* **1966**, *9*, 60–64. https://doi.org/10.1093/comjnl/9.1.60.
11. Hassanat, A.B. Dimensionality Invariant Similarity Measure. *J. Am. Sci.* **2014**, *10*, 221–226. https://doi.org/10.7537/marsjas100814.31.
12. Batista, G.E.A.P.A.; Keogh, E.J.; Tataw, O.M.; de Souza, V.M.A. CID: An efficient complexity-invariant distance for time series. *Data Min. Knowl. Discov.* **2014**, *28*, 634–669. https://doi.org/10.1007/s10618-013-0312-3.
13. Abu Alfeilat, H.A.; Hassanat, A.B.A.; Lasassmeh, O.; Tarawneh, A.S.; Alhasanat, M.B.; Eyal Salman, H.S.; Prasath, V.B.S. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data* **2019**, *7*, 221–248. https://doi.org/10.1089/big.2018.0175.
14. Bagnall, A.; Lines, J.; Bostrom, A.; Large, J.; Keogh, E. The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Discov.* **2017**, *31*, 606–660. https://doi.org/10.1007/s10618-016-0483-9.
15. Bogatinovski, J.; Todorovski, L.; Džeroski, S.; Kocev, D. Comprehensive Comparative Study of Multi-Label Classification Methods. *Expert Syst. Appl.* **2022**, *203*, 117215. https://doi.org/10.1016/j.eswa.2022.117215.
16. Du, S.; Zhang, X.; Lei, Y.; Huang, X.; Tu, W.; Liu, B.; Meng, Q.; Du, S. Mapping urban functional zones with remote sensing and geospatial big data: a systematic review. *GIScience Remote Sens.* **2024**, *61*, 2404900. https://doi.org/10.1080/15481603.2024.2404900.
17. Soto, V.; Frías-Martínez, E. Automated Land Use Identification using Cell-Phone Records. In *Proceedings of the 3rd ACM International Workshop on Hot Topics in Planet-Scale Measurement, HotPlanet'11*; ACM: New York, NY, USA, 2011. https://doi.org/10.1145/2000172.2000179.
18. Blondel, V.D.; Decuyper, A.; Krings, G. A survey of results on mobile phone datasets analysis. *EPJ Data Sci.* **2015**, *4*, 10. https://doi.org/10.1140/epjds/s13688-015-0046-0.
19. De Nadai, M.; Staiano, J.; Larcher, R.; Sebe, N.; Quercia, D.; Lepri, B. The Death and Life of Great Italian Cities: A Mobile Phone Data Perspective. In Proceedings of the 25th International Conference on World Wide Web (WWW), Montreal, QC, Canada, 11–15 May 2016; pp. 413–423. https://doi.org/10.1145/2872427.2883084.
20. Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and POIs. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 186–194. https://doi.org/10.1145/2339530.2339561.
21. Mendoza-Hurtado, M.; Romero-del Castillo, J.A.; Ortiz-Boyer, D. SAMPLID: A New Supervised Approach for Meaningful Place Identification Using Call Detail Records as an Alternative to Classical Unsupervised Clustering Techniques. *ISPRS Int. J.-Geo-Inf.* **2024**, *13*, 289. https://doi.org/10.3390/ijgi13080289.

22. Oosterlinck, D.; Baecke, P.; Benoit, D.F. Home location prediction with telecom data: Benchmarking heuristics with a predictive modelling approach. *Expert Syst. Appl.* **2021**, *170*, 114507. https://doi.org/10.1016/j.eswa.2020.114507.
23. Verma, R.; Mittal, S.; Lei, Z.; Chen, X.; Ukkusuri, S.V. Comparison of home detection algorithms using smartphone GPS data. *EPJ Data Science* **2024**, *13*, 6. https://doi.org/10.1140/epjds/s13688-023-00447-w.
24. Madjarov, G.; Kocev, D.; Gjorgjevikj, D.; Džeroski, S. An extensive experimental comparison of methods for multi-label learning. *Pattern Recognit.* **2012**, *45*, 3084–3104. https://doi.org/10.1016/j.patcog.2012.03.004.
25. Romero-Del-Castillo, J.; Mendoza-Hurtado, M.; Ortiz-Boyer, D.; García-Pedrajas, N. Local-based k values for multi-label k-nearest neighbors rule. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105487. https://doi.org/10.1016/j.engappai.2022.105487.
26. Cuevas-Muñoz, J.M.; García-Pedrajas, N. ML-k'sNN: Label Dependent *k* Values for Multi-Label *k*-Nearest Neighbor Rule. *Mathematics* **2023**, *11*, 275. https://doi.org/10.3390/math11020275.
27. Gouk, H.; Pfahringer, B.; Cree, M. Learning Distance Metrics for Multi-Label Classification. In *Proceedings of the 8th Asian Conference on Machine Learning (ACML)*; Proceedings of Machine Learning Research: Cambridge, MA, USA, 2016; Volume 63. https://doi.org/10.70675/1dfcd0cfzb7ebz4bfcza95az2594c697e708.
28. Mao, J.X.; Wang, W.; Zhang, M.L. Label Specific Multi-Semantics Metric Learning for Multi-Label Classification: Global Consideration Helps. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI), Macao, China, 19–25 August 2023. https://doi.org/10.24963/ijcai.2023/451.
29. Mao, J.X.; Hang, J.Y.; Zhang, M.L. Learning Label-Specific Multiple Local Metrics for Multi-Label Classification. In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI), Jeju, Republic of Korea, 3–9 August 2024; pp. 4742–4750. https://doi.org/10.24963/ijcai.2024/524.
30. Ding, H.; Trajcevski, G.; Scheuermann, P.; Wang, X.; Keogh, E. Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proc. VLDB Endow.* **2008**, *1*, 1542–1552. https://doi.org/10.14778/1454159.1454226.
31. Paparrizos, J.; Li, H.; Yang, F.; Wu, K.; d'Hondt, J.E.; Papapetrou, O. A Survey on Time-Series Distance Measures. *arXiv* **2024**, arXiv:2412.20574. https://doi.org/10.48550/arXiv.2412.20574.
32. Ruiz, A.P.; Flynn, M.; Large, J.; Middlehurst, M.; Bagnall, A. The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Min. Knowl. Discov.* **2021**, *35*, 401–449. https://doi.org/10.1007/s10618-020-00727-3.
33. d'Hondt, J.E.; Li, H.; Yang, F.; Papapetrou, O.; Paparrizos, J. A Structured Study of Multivariate Time-Series Distance Measures. *ACM Manag. Data (SIGMOD)* **2025**, *3*, 121. https://doi.org/10.1145/3725258.
34. Cuturi, M.; Blondel, M. Soft-DTW: A Differentiable Loss Function for Time-Series. In *Proceedings of the 34th International Conference on Machine Learning*; Proceedings of Machine Learning Research (PMLR): Cambridge, MA, USA, 2017, Volume 70, pp. 894–903.
35. Sechidis, K.; Tsoumakas, G.; Vlahavas, I. On the Stratification of Multi-label Data. In *Proceedings of the Machine Learning and Knowledge Discovery in Databases*; Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 145–158. https://doi.org/10.1007/978-3-642-23808-6_10.
36. Szymański, P.; Kajdanowicz, T. A Network Perspective on Stratification of Multi-Label Data. In *Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications*; Proceedings of Machine Learning Research (PMLR): Cambridge, MA, USA, 2017; Volume 74, pp. 22–35.
37. Kattenborn, T.; Schiefer, F.; Frey, J.; Feilhauer, H.; Mahecha, M.D.; Dormann, C.F. Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. *ISPRS Open J. Photogramm. Remote Sens.* **2022**, *5*, 100018. https://doi.org/10.1016/j.ophoto.2022.100018.
38. Karasiak, N.; Dejoux, J.F.; Monteil, C.; Sheeren, D. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Mach. Learn.* **2022**, *111*, 2715–2740. https://doi.org/10.1007/s10994-021-05972-1.
39. Wang, Y.; Khodadadzadeh, M.; Zurita-Milla, R. A new cross-validation method to evaluate geospatial machine learning models. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *121*, 103364. https://doi.org/10.1016/j.jag.2023.103364.
40. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech, Signal Process.* **1978**, *26*, 43–49. https://doi.org/10.1109/TASSP.1978.1163055.
41. Provincia Autonoma di Trento. ISTAT: In Trentino Bene l'occupazione, Tiene il PIL. Ufficio Stampa Della Provincia Autonoma di Trento. 2015. Available online: https://www.ufficiostampa.provincia.tn.it/Comunicati/ISTAT-IN-TRENTINO-BENE-L-OCCUPAZIONE-TIENE-IL-PIL (accessed 7 November 2026).
42. Provincia Autonoma di Trento. Dati e Statistiche. Official Provincial Data Portal. Available online: https://www.provincia.tn.it/Dati-e-statistiche (accessed 30 September 2026).
43. Juszczak, P.; Tax, D.M.J.; Duin, R.P.W. Feature Scaling in Support Vector Data Description. In Proceedings of the 8th Annual Conference of the Advanced School for Computing and Imaging (ASCI), Delft, The Netherlands, 19–21 June 2002; pp. 95–102.
44. Tsoumakas, G.; Katakis, I. Multi-label classification: An overview. *Int. J. Data Warehous. Min.* **2007**, *3*, 1–13. https://doi.org/10.4018/jdwm.2007070101.

45. Wang, D.; Wang, J.; Hu, F.; Li, L.; Zhang, X. A Locally Adaptive Multi-Label k-Nearest Neighbor Algorithm. In *Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 81–93. https://doi.org/10.1007/978-3-319-93034-3_7.
46. Read, J.; Pfahringer, B.; Holmes, G.; Frank, E. Classifier chains for multi-label classification. *Mach. Learn.* **2011**, *85*, 333–359. https://doi.org/10.1007/s10994-011-5256-5.
47. Tsoumakas, G.; Vlahavas, I. Random k-Labelsets: An Ensemble Method for Multilabel Classification. In *Proceedings of the European Conference on Machine Learning (ECML)*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 406–417. https://doi.org/10.1007/978-3-540-74958-5_38.
48. Cawley, G.C.; Talbot, N.L.C. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.