## Check Status

**Submission ID:** 277

**Title:** CommuniWave 2.0--A Spatially-Enhanced Machine Learning Model for Quantifying the Degree of Temporary Informal Behavior in Urban Communities

**Status:** Accept

**Reviewer 1:**

***Author Comments:*** Application of the research in practice or academic is unclear and needs to be clarified. Limitations need to address the boundaries of the context. Also the skills required to adopt these processes need to be further acknowledged.

**Reviewer 2:**

***Author Comments:*** The manuscript presents a multi-modal pipeline (behaviour recognition, monocular depth, semantic segmentation, entropy-based spatial analysis) to detect and score dynamic informal behaviours for governance support. The framing is relevant and the pipeline design is technically credible. To enhance rigour, report dataset composition, annotation protocols, inter-rater agreement, and privacy safeguards; present cross-city generalisation, ablation studies for each spatial component, error analysis in dense scenes, and calibration between model scores and stakeholder-defined thresholds. Clarify intended decision workflows and risks of algorithmic bias, with mitigation strategies and uncertainty communication.

**Reviewer 3:**

***Author Comments:*** The dataset is drawn entirely from a single mid-sized city in Southern China, yet the paper claims broad portability. Informal behaviour varies greatly across cultural, morphological, and governance contexts, which may limit the model’s transferability. The discussion should explicitly address potential context-specific bias and outline how CW2.0 could be adapted (via transfer learning, re-weighting of spatial features, or recalibration of scoring criteria) for different urban morphologies and regulatory models.

The SCM produces absolute metric estimates (distances, m^2 areas) from monocular depth. Given the importance of proximity thresholds (1 m / 2 m) and area shares in the final prediction, the absence of validation of metric accuracy is quite a gap. Please reveal expected or observed reconstruction errors for the given camera geometry, or reference ZoeDepth accuracy under comparable configurations.

DIB labels average the perceptions of city managers, residents, and vendors, but this may conceal meaningful governance conflicts (e.g., control vs. vibrancy). The study should discuss how averaging affects interpretability and consider alternatives such as predicting stakeholder-specific DIB scores or analysing residuals to identify systematic bias linked to particular groups.

The integrated detector–depth–segmentation–entropy pipeline is computationally heavy. Since the manuscript emphasises real-time or routine monitoring applications, inference speed is critical. Please provide end-to-end FPS for the full CW2.0 pipeline on the stated hardware and discuss optimisation strategies (e.g., pruning, quantisation) needed for practical deployment.

The combined feature set improves accuracy but reduces transparency. For an urban design audience, it is important to understand which spatial factors contribute most to DIB scores.

Including XGBoost feature-importance analysis (e.g., permutation or split-gain) would provide design-oriented insight and strengthen the link between the technical method and spatial reasoning.

The clarity and readability of Figures 1 and 2 (workflow diagrams) are insufficient. Text within the modules is blurry, and key labels for components such as BCN_E, SCM submodules, and XGBoost are difficult to read.

The comparison of Figure 4 is not well standardised. The CW1.0 and CW2.0 outputs use different crops and framing, which weakens the credibility of the comparison and makes the result feel illustrative rather than evaluative. To clearly demonstrate improvement, both versions should use identical viewpoints and explicitly point out how CW2.0 resolves specific issues seen in CW1.0, such as occlusion or clustering errors.

---