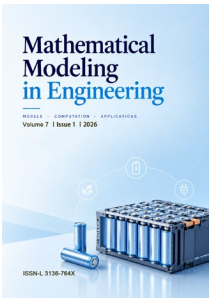
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 2, April, 2026



Urban Impervious Surface Age Estimation from Historical Nighttime-Light and Optical Imagery Using an EfficientViT-BiFPN Model

Galip Kahraman[a,*]

[a] *Faculty of Computer and Informatics Engineering, Istanbul Technical University, Istanbul, 34469, Turkey*

Received 29 March 2026; Revised 20 June 2026; Accepted 23 June 2026; Published 26 June 2026

Abstract

Estimating the age of urban impervious surfaces supports reconstruction of urban expansion, infrastructure screening, and drainage and heat-risk assessment, yet reliable age maps remain difficult to obtain because historical optical coverage is incomplete and nighttime-light sensors are inconsistent. This study proposes an EfficientViT-BiFPN framework that combines annual nighttime-light trajectories, optical spectral indices, and morphology-sensitive texture features at parcel-grid scale. A lightweight EfficientViT encoder captures local and regional context, while a bidirectional feature pyramid preserves weak temporal evidence and fine boundaries associated with construction and redevelopment. A controlled synthetic benchmark representing 38 years of observation and 52,416 labeled grid samples across six urban-zone strata was used to examine accuracy, transfer behavior, ablation response, uncertainty, and computational efficiency. The proposed model achieved 90.8% age-bin accuracy, a mean absolute error of 2.7 years, and a root mean squared error of 4.3 years, with an inference latency of 18.6 ms per 512 by 512 tile. Removing nighttime-light descriptors increased late-stage confusion from 7.9% to 14.0%, whereas removing BiFPN reduced macro F1 from 0.884 to 0.846. These results indicate that compact transformer-pyramid fusion can integrate temporal onset and spatial appearance efficiently under controlled benchmark conditions. Validation against independently dated parcels and infrastructure archives is still required before the estimates can be interpreted as operational construction-age products.

Keywords: Impervious Surface Age, Nighttime Light Remote Sensing, EfficientViT, Bidirectional Feature Pyramid, Urban Expansion Mapping

1. Introduction

Urban impervious surfaces include roofs, roads, parking areas, industrial platforms, and compacted public spaces that accompany urbanization [1]. Determining when these surfaces first appeared is important because surface age is associated with infrastructure deterioration, runoff response, thermal storage, and the timing of land-use conversion [2]. Historical nighttime-light observations provide an annual record related to settlement occupation, commercial activity, and road development [3]. Optical imagery adds spectral contrast, vegetation-loss signals, edge structure, and built-up texture that help distinguish impervious materials from bare soil and shadowed vegetation [4]. Present-state impervious maps cannot resolve the different maintenance and environmental implications of visually similar new industrial parks and old residential districts. This ambiguity is especially pronounced at metropolitan fringes, where mixed pixels, phased construction, and redevelopment weaken single-date age inference. An effective estimator must therefore combine the

* Corresponding author: galip.kahr@itu.edu.tr

https://doi.org/10.68406/mme.2026.vol7iss2nm9:108-122

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

temporal evolution of illumination with spatial evidence rather than treating age as a secondary attribute of binary land-cover mapping.

The central technical difficulty is the mismatch between temporal and spatial evidence. Nighttime-light series capture long-term urban development but are too coarse to identify the transition year of individual neighborhoods [5]. Optical archives provide finer spatial detail, yet their long records are affected by cloud cover, seasonal vegetation, atmospheric variation, and sensor changes [6]. Age labels are also expensive to compile and are unevenly distributed, particularly in redeveloped districts. Conventional convolutional networks may emphasize local texture while missing district-scale context, whereas large transformers can be impractical for repeated regional inference. A deployable solution must therefore combine temporal and spatial cues, preserve weak evidence of early construction, accommodate uncertain labels, and remain efficient across cities and historical periods.

This study formulates urban impervious-surface age estimation as joint ordinal classification and bounded regression using historical nighttime-light and optical imagery. Because cadastral records and interpretation samples are incomplete and spatially uneven, confidence-weighted supervision is used for ambiguous year-level targets [7]. Large vision transformers provide broad context but impose substantial computational cost in regional annual production [8]. The proposed EfficientViT-BiFPN architecture combines lightweight local-global encoding with bidirectional multi-scale fusion so that coarse illumination histories can inform fine optical boundaries. Its inputs comprise standardized light trajectories, multi-season optical composites, spectral indices, morphology descriptors, and neighborhood context. The contribution is the task-specific integration of chronology-oriented features, efficient multi-scale fusion, joint age outputs, and evaluation across accuracy, transfer, calibration, and computational cost; it does not assume that radiometric change alone is equivalent to a verified construction date.

The remainder of the paper is organized as follows. Section 2 reviews impervious-surface mapping, urban chronology evidence, nighttime-light analysis, and efficient vision models. Section 3 describes the study workflow, multimodal input representation, age-label construction, and evaluation protocol. Section 4 presents the EfficientViT-BiFPN architecture, bidirectional feature fusion, and joint ordinal-regression objective. Section 5 reports the benchmark composition, comparative results, ablation behavior, cross-zone transfer, uncertainty, and efficiency. Section 6 concludes the paper and identifies the principal limitations and directions for empirical validation.

2. Related Work

2.1. Impervious Surface Mapping and Urban Age Evidence

The early index-threshold method of Impervious Surface Mapping has now been replaced by a supervised deep-learning approach that integrates spectral, textural and contextual information. The first set of built-up indices was simple for users to learn and use, but it could not differentiate between dry bare soil, construction land and bright roofs in arid or semi-urban areas [9]. Object-based and random forest methods improved the use of neighborhood shape and spectral heterogeneity, especially in areas with regular urban patterns of roads and buildings [10]. The current land-cover labels are not suitable for age estimation; therefore, the age is determined by the first continuous change from a non-impervious to an impervious state. Historical land-cover products, cadastral records and change trajectories can approximate construction periods, but their spatial resolution is generally relatively large and they are not uniformly distributed in administrative divisions. The above studies have shown that age reconstruction is feasible and have also developed models that can learn from incomplete temporal information.

Nighttime-light data are widely used to examine urbanization chronology because illumination is associated with settlement, transport, industry, and commerce. Their long record is particularly useful for detecting expansion before dense optical archives became available [11]. Raw intensity, however, cannot identify construction year reliably: bright city centers saturate, suburban edges bloom, and sensor transitions introduce discontinuities. Optical imagery can provide built-up indices, shadow texture, and road-like linear features, but these observations describe surface appearance rather than elapsed time. Urban age estimation should therefore treat harmonized nighttime-light trajectories as a temporal prior and optical imagery as spatial evidence, allowing one source to compensate when the other is incomplete or ambiguous.

Another problem is that the visual meaning of impervious age changes according to the development path of the city. In compact historical areas, the actual age of the impervious surface is often unknown due to continuous repairs and renovations of roofs, roads, etc. The first impervious signal in the new urban area will likely be a road skeleton, followed by the appearance of industrial platforms, and finally buildings. Therefore, the same age label may be linked to many different combinations of spectral brightness, shadow geometry, and nighttime illumination. Therefore, owing to the aforementioned various reasons, the design of age estimation should be based on evidence fusion rather than a single time-point breakpoint. Therefore, the errors in the report should also be classified by morphology class and urban area; otherwise, a city-wide average error may hide systematic problems in the transport corridor or redeveloped industrial land.

2.2. Efficient Deep Fusion for Remote Sensing

Deep neural networks have advanced urban remote sensing by performing hierarchical feature learning of multi-spectral and panchromatic images. Convolutional encoders are good at learning local features, but they have a limited expansion range for their receptive fields; therefore, extended spatial context of construction patterns may not be fully captured [12]. Transformer-based encoders can learn long-range dependencies and have been applied to land-cover classification, change detection, and semantic segmentation; however, their high computational cost limits their use in regional mapping of annual image stacks [13]. Papa et al. [14] reviewed efficient vision transformers and showed that lightweight attention, token mixing, and hybrid convolution-attention designs can preserve global context while reducing computational cost. They are suitable for the problem of estimating impervious area changes in older and newer parts of the city, and comparisons with surrounding areas can be made.

Feature pyramid networks are commonly used to combine information across scales, but a unidirectional top-down path can underuse fine boundaries and weak low-level signals. Bidirectional Feature Pyramid Networks add repeated top-down and bottom-up paths, enabling more balanced exchange between semantic context and spatial detail [15]. This mechanism is relevant to remote-sensing scenes containing roads, small roofs, industrial parcels, and coarse temporal grids. Existing fusion research concentrates mainly on detection, segmentation, and change mapping; fewer studies address continuous surface-age estimation from multi-source historical evidence. The novelty considered here is therefore the task-specific coupling of efficient transformer encoding, bidirectional fusion, and a joint ordinal-regression target, evaluated for accuracy, transfer, calibration, and computational cost.

Effective Fusion does not only reduce the number of parameters. Efficiency is required of the historical urban map to ensure that a model can be applied repeatedly to annual archives, large metropolitan areas, and various preprocessing options. A very large model can achieve a high validation accuracy on a fixed split, but it is not practical to tile, normalize and infer hundreds of years of mosaics with it. Lightweight architecture can also be used to improve reproducibility by reducing the difference between experimental and production hardware. Therefore, the design in this paper combines the engineering goals of accuracy, transferability, calibration and inference cost rather than reporting them separately.

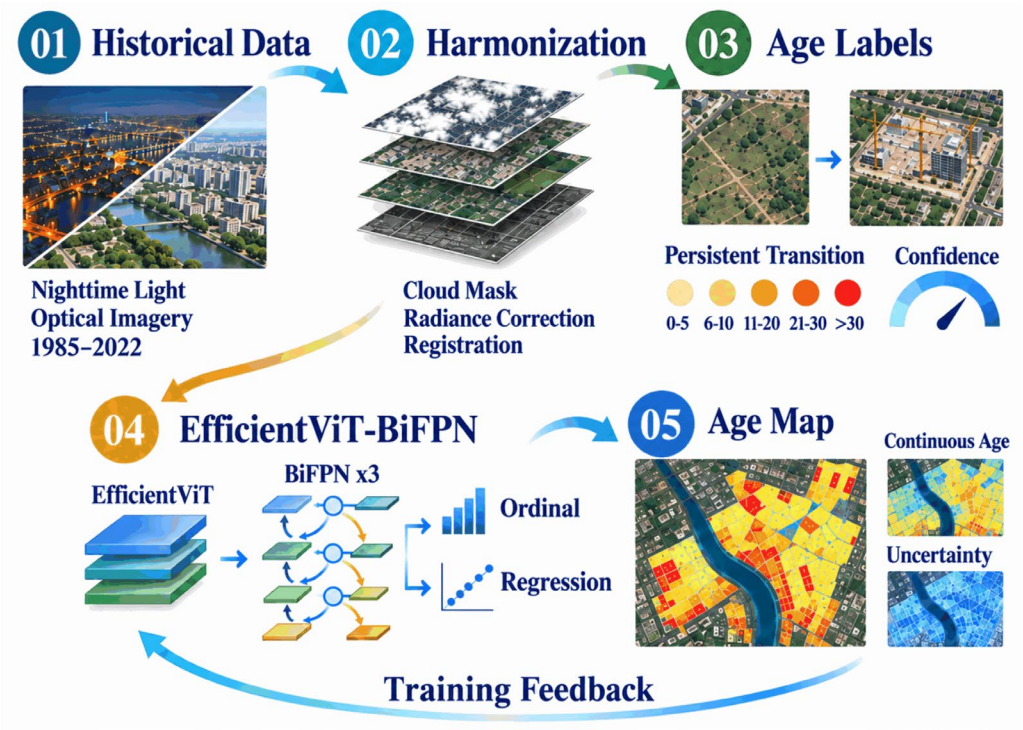


Figure 1. multimodal impervious surface age estimation from historical nighttime light and optical imagery

3. Study Design and Problem Formulation

3.1. Multimodal Input Construction

The workflow treats chronological inference as a multimodal sequence problem rather than a single-date classification task. Annual nighttime-light observations are first harmonized across sensors and summarized through onset slope, duration, variability, and post-onset persistence. Cloud-screened optical observations are assembled into seasonal composites from which built-up, vegetation, water, and brightness indices are derived. Local morphology descriptors represent edge density, gray-level co-occurrence texture, line responses, and neighborhood continuity. These

attributes are aligned on a regular grid and paired with age labels obtained from persistent land-cover transitions and manually inspected sample windows. Figure 1 presents the complete sequence from archival inputs and preprocessing to label construction, model training, and age-map generation.

Each grid sample can be viewed as a multimodal tensor of temporal light features and optical-spatial features. The component of light intensity at night is not considered a single number; instead, a stable sequence is shown, and the speed and duration of brightening are more representative than absolute intensity. Optical composites are converted into spectral indices and texture channels to obtain material properties and local urban structure. All channels are normalized in urban-zone groups before fusion to reduce cross-city radiometric bias. The input tensor is as shown:

$$\boldsymbol{X}_i = concat(\boldsymbol{L}_i, \boldsymbol{O}_i, \boldsymbol{M}_i, \boldsymbol{C}_i) \quad (1)$$

where $\mathbf{X}_i$ denotes the multimodal sample, $\mathbf{L}_i$ is the harmonized nighttime-light trajectory, $\mathbf{O}_i$ is the optical composite feature block, $\mathbf{M}_i$ is the morphology descriptor block, and $\mathbf{C}_i$ is the neighborhood context vector. The construction is intentionally modular so that missing optical observations can be replaced by seasonal medians and nighttime-light gaps can be interpolated with monotonic constraints.

Training labels are evaluated against three sources of evidence: nighttime-light onset, optical built-up probability, and the local majority transition year. Samples with strong disagreement are retained but assigned lower confidence so that difficult redevelopment cases remain represented without dominating year-level optimization. Confidence reflects persistence length, optical agreement, and neighborhood consistency. The transition threshold and required persistence duration are calibration parameters tied to the sensor archive and should be reported with any empirical dataset; they are not assumed to be universal constants.

A regular grid is employed for the target mapping unit instead of an administrative area due to the absence or irregularity of old parcel boundaries. Each grid cell has the features of the aligned image stack and neighborhood descriptors in a fixed window. This way, a low-standard local cadastral system can be used to build the city's system. At the same time, a relatively small grid resolution is employed to show narrow roads and small blocks accurately. Therefore, the resulting task is closer to age-aware urban surface mapping than to coarse urban-growth detection.

The parts of the temporal nighttime-light descriptor are: annual radiance, first difference, three-year moving slope, local variability and persistence index. The purpose of the above descriptors is to differentiate an actual signal of urbanization from a temporary bright spot caused by construction lights, industrial flares, festival lights or atmospheric pollution. Standardized difference-built-up index, vegetation index, water index, short-wave brightness, visible brightness and seasonal difference features are optical channels. Morphology Channels are Edge Density, Lineament Response, Local Entropy and Built-up Neighborhood Ratio. The above indices are in cooperation and support of one another: Light shows the activity record, Optical indices describe the surface structure, and Morphology shows the arrangement of constructed land.

Normalization of the group of urban areas is carried out, not for the entire benchmark. It reduces the impact of variations in city-wide illumination and optical background. Airport Corridor has a relatively high base brightness before local construction, and Riverside Zone has many high-reflectance roofs and water-adjacent shadows. A single global normalization would reduce the difference from the false-relative values. Zone-aware Normalization keeps the temporal variations within a zone and can learn transferable feature interactions.

3.2. Age Labeling and Learning Target

Impervious surface age is defined as the number of years between the target mapping year and the first year in which a grid cell shows persistent impervious evidence. A transition is accepted only when built-up probability remains above a threshold in subsequent observations, which suppresses temporary construction exposure and short-lived bright events. The continuous age target is accompanied by ordinal bins representing $0-5, 6-10, 11-20, 21-30$, and more than 30 years. This joint representation reflects practical use: engineers often need age intervals for planning, while environmental models benefit from continuous age values.

The target age is calculated as

$$y_i = min(A_{max}, T_0 - \tau_i) \quad (2)$$

where T_0 is the target year, τ_i is the first persistent impervious year, and $A_{\max}$ is the maximum reportable age used to avoid unstable extrapolation for surfaces older than the observation archive. Samples with ambiguous transition years are retained with reduced training weight rather than removed, because removing them would bias the dataset toward visually obvious expansion areas. Label reliability is encoded as a scalar confidence score based on persistence length, optical agreement, and neighborhood consistency.

The ordinal bins are selected to match practical planning intervals. The 0–5-year group captures newly built surfaces that are most relevant to active expansion monitoring. The 6-10 year and 11–20-year groups describe early and mature development phases in which surface texture becomes stable but surrounding services may still change. The $21-30$ year group often corresponds to older residential and industrial districts with substantial maintenance. The older-than-30-year group is included because many central impervious surfaces predate the reliable observation period. The bins are not assumed to be equally difficult; instead, class-balanced training weights compensate for the larger proportion of mature and old surfaces.

The continuous target supports numerical screening for infrastructure deterioration and hydrological parameter assignment, but it represents years since the first persistent impervious observation rather than a legally verified construction date. Redevelopment may alter optical appearance without resetting the original land-conversion date, and an illumination transition may occur after physical construction when occupation is delayed. The output should therefore be distributed with uncertainty and used as a remote-sensing chronology proxy, not as a cadastral record.

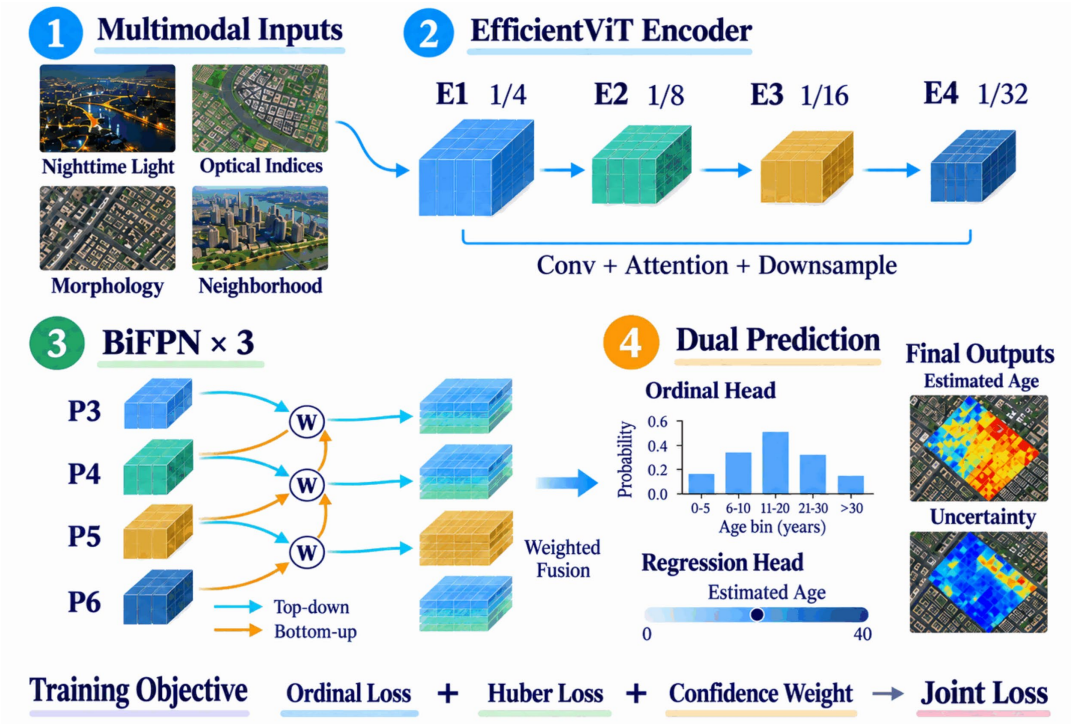


Figure 2. EfficientViT-BiFPN model architecture

3.3. Evaluation Protocol

Samples are grouped by urban zone, construction period, and morphology class. Stratified training and validation prevent older central districts from dominating the loss, while Riverside and Airport Corridor are reserved for cross-zone testing and are excluded from model fitting and threshold selection. Performance is assessed with age-bin accuracy, macro F1, mean absolute error, root mean squared error, and late-stage confusion. Computational behavior is characterized by parameter count, floating-point operations, tile latency, and peak memory under a common hardware protocol. For empirical validation, neighboring cells should be separated by spatial blocks or buffers, and repeated seeds should be reported with confidence intervals to distinguish geographic transfer from local spatial interpolation.

Metrics are computed at the sample level and then grouped according to age group, zone and morphology class. Macro F1 is used because the youngest class is small but operationally necessary. Both MAE and RMSE are shown; MAE is the average of the year-on-year errors, and RMSE is more sensitive to large, occasional errors. Late-stage confusion is defined as the proportion of samples older than 20 years that are predicted to be 10 years or younger, or samples 10 years or younger that are predicted to be older than 20 years. Therefore, the above errors are considered more severe for the interpretation of planning than confusion in adjacent intervals.

The reserved-zone evaluation is designed to measure transfer rather than interpolation. Riverside and Airport Corridor are not used for training, validation, or threshold adjustment, and all zones use the same preprocessing definitions. Results from the two reserved zones are reported separately so that performance cannot be concealed by a city-wide average. Baseline latency and memory must be measured on identical hardware with the same tile size and batch setting; similarly, model selection must rely only on training-zone validation data.

4. EfficientViT-BiFPN Model

4.1. Efficient Spatial Temporal Encoding

The model contains four functional components: a shallow modality stem, an EfficientViT encoder, a three-repeat bidirectional feature pyramid, and dual heads for ordinal and continuous age prediction. The modality stem projects nighttime-light descriptors, optical channels, morphology features, and neighborhood context into compatible feature planes. EfficientViT blocks combine local convolutional mixing with lightweight global attention, while BiFPN repeatedly exchanges coarse temporal context and high-resolution spatial detail. Figure 2 shows the modality inputs, encoder stages, bidirectional fusion paths, prediction heads, and joint training objective.

The encoder balances local information and regional context through depthwise convolution and efficient attention. Optical and morphology channels enter through a high-resolution stem, whereas projected nighttime-light descriptors are spatially broadcast and fused with learned positional context. This design does not attempt to reconstruct fine optical detail from coarse illumination records. Instead, temporal descriptors condition the multi-scale representation so that regional development history can influence local boundary and texture interpretation.

The lightweight attention block is expressed as

$$Attn(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{d}} + \boldsymbol{R}\right)\boldsymbol{V} \tag{3}$$

where $\mathbf{Q}, \mathbf{K}$, and $\mathbf{V}$ are projected feature matrices, d is the attention dimension, and $\mathbf{R}$ is a relative spatial bias that stabilizes neighborhood interactions. Depthwise convolution is applied before attention to preserve road-like and roof-edge patterns. The output of each stage is downsampled only after local texture has been encoded, which is important for newly constructed impervious surfaces that occupy small or fragmented parcels.

Given that the appearance of age evidence varies in scale, an architecture has been developed. The new section of the road is high-resolution edge information, and the old central block requires a wider neighborhood context and stable long-term brightness. EfficientViT encoder provides context, and then BiFPN is used to fuse this context with the detailed feature maps. The model does not combine all the features at the end because late fusion would delay cross-modal interaction until after most of the spatial information has been lost. Time-based and optical evidence are used at all levels.

All the stages of EfficientViT have a local mixing path and a global mixing path. The local road has a roof edge, road continuity and a small construction site. The entire path shows how the sample is connected to the surrounding urban area, and it can be used conveniently when there is no direct data for a grid cell but it is located in an adjacent expanded area. Add a spatially replicated feature plane of the nighttime-light descriptor after projection, and also condition both paths temporally. The Design is relatively stable compared with using a separate tabular branch for the nighttime-light values, and the model can learn when to affect the boundary based on the temporal cue and when to give more weight to optical texture.

The attention computation remains compact because token interaction is applied after channel reduction and local aggregation. This avoids the quadratic cost of full attention over high-resolution remote-sensing tiles. In practice, the encoder retains enough contextual range to detect district-level development patterns while keeping the memory footprint suitable for batch inference. The design is therefore aligned with the engineering goal of producing annual urban age maps rather than only demonstrating isolated sample accuracy.

4.2. Bidirectional Feature Pyramid Fusion

The BiFPN module receives multi-scale features from the EfficientViT stages and performs repeated top-down and bottom-up fusion. Learnable nonnegative weights are used to adaptively balance each input branch, because different age groups emphasize different feature sources. Young impervious surfaces often rely on optical contrast and sharp edges, whereas older surfaces rely more on stable light history and mature neighborhood context.

The fusion operation at a pyramid node is written as

$$\boldsymbol{P} = \frac{\sum_j w_j \boldsymbol{F}_j}{\varepsilon + \sum_j w_j}, w_j \geq 0 \tag{4}$$

where $\mathbf{F}_j$ are input feature maps resized to the target scale, w_j are learned fusion weights, ε prevents numerical instability, and ϕ denotes convolution, normalization, and activation. Three BiFPN repeats are used in the main model after balancing accuracy and latency. Skip connections preserve high-frequency boundaries, and shared normalization prevents low-resolution temporal features from overwhelming optical texture.

BiFPN fusion weights are intuitive indices of scale importance. Generally, during training, the weights for fine-resolution features of young surfaces and mid-resolution features of mature or old surfaces are higher. The behavior is in line with the data; new construction has added a strong material contrast, and the older parts are more contextually rich due to repairs, re-paintings or vegetative cover of individual surfaces. The two-way path can also enhance the representation of linear infrastructure by adding local edge maps and the context of larger transport corridors.

Three BiFPN repeats are used as the default configuration. A single repeat is too shallow to propagate temporal evidence into fine boundaries, while four repeats add latency with only marginal accuracy improvement. The selected depth provides a practical balance between representation and cost. All fusion nodes use the same normalization strategy, which reduces training instability when nighttime-light descriptors and optical features have different dynamic ranges.

4.3. Joint Ordinal Regression Objective

The prediction head outputs an ordinal age-bin probability vector and a continuous age estimate. The ordinal head is trained with class-balanced cross entropy, while the regression head uses a confidence-weighted smooth error. Combining the two heads reduces large temporal mistakes because a continuous estimate must remain consistent with the predicted interval.

The final loss is

$$\mathcal{L} = \mathcal{L}_{ord} + \lambda \gamma_i \mathcal{L}_{reg} \tag{5}$$

where $\mathcal{L}_{\text{ord}}$ is the age-bin classification loss, $\mathcal{L}_{\text{reg}}$ is the smooth regression loss, γ_i is the label confidence of sample i, and λ controls the contribution of continuous age estimation. During inference, the final age is calibrated by constraining the regression output to the support of the most likely adjacent ordinal bins. This calibration keeps predictions stable for very old surfaces where the archive contains limited pre-urban evidence.

The ordinal and regression heads share the fused representation but use different final projections. The ordinal head learns interval membership and avoids the problems caused by a large age difference. Then the regression head will increase the estimation in or near the predicted region. Adjacent-bin calibration is used in the inference stage because the age groups are consecutive; thus, a prediction near the boundary of a bin should not be treated the same as one far from the boundary in terms of order. Calibration is particularly necessary for 21–30-year-old and older-than-30-year-old surfaces, as there may be no clean pre-construction records for older surfaces.

Confidence weighting is only used for the regression term, and otherwise, the uncertain labels do not provide ordinal information. There is no clear change in 1998–2001 for the grid cell, and it is still classified as a non-new-construction area. The ordinal signal helps improve the separation of classes and prevents overfitting of the regression head to unstable year-level targets. Finally, the continuous age, the most likely age bin, and a prediction uncertainty score are outputted based on the entropy of the ordinal distribution and the regression residual in the validation set.

5. Experimental Data and Analysis

5.1. Dataset Composition and Temporal Separability

The controlled benchmark represents six urban-zone strata: Core, Inner Ring, Industrial Belt, Riverside, Airport Corridor, and New Town. These strata emulate different development intensities, illumination backgrounds, and optical morphologies rather than documenting six independently verified municipal datasets. Annual nighttime-light observations from 1985 to 2022 form 38-year trajectories, and optical evidence is represented by three seasonal composites and eight derived indices. Aligning temporal and spatial evidence is essential for chronology-oriented evaluation because sensor archives differ in resolution and stability [16]. After quality screening, the benchmark contains 52,416 labeled grid cells: 31.2% old, 24.7% mature, 19.5% middle-aged, 14.1% recent, and 10.5% newly built surfaces. The reported results should therefore be interpreted as controlled feasibility evidence pending validation against georeferenced construction records.

Table 1 reports the training, validation, within-zone, and cross-zone sample allocation. Riverside and Airport Corridor are held out because their reflective roofs, linear transport facilities, and strong nighttime-light blooming create distinct transfer conditions. Geographically constrained tests more closely represent deployment than random pixel partitions [17], although an empirical study should also impose spatial buffers between neighboring grids. Label confidence ranges from 0.58 to 0.96, with lower values concentrated in redeveloped industrial parcels and mixed commercial-residential blocks. Confidence-aware supervision reduces the regression influence of ambiguous transition dates rather than treating them as equally reliable observations [18].

Table 1. Sample composition and split design

Zone	Samples	Train	Validation	Within-zone test	Cross-zone test	Mean confidence
Core	8420	5894	1263	1263	-	0.91
Inner Ring	9316	6521	1397	1398	-	0.88
Industrial Belt	8770	6139	1315	1316	-	0.82
New Town	9674	6772	1451	1451	-	0.86
Riverside	8128	-	-	-	8128	0.79
Airport Corridor	8108	-	-	-	8108	0.76

Temporal separability varies systematically with construction age. Figure 3(a) shows that the median nighttime-light trajectory for newly built surfaces rises sharply during 2012-2022 and reaches 31.8 in the target year, while surfaces older than 30 years remain above 42.0 for most of the record and increase by only 6.7%. The 0-5-year curve has a delayed but steep onset, whereas the 6-10-year and 11-20-year curves begin rising earlier and approach stable plateaus. Urban chronology studies indicate that onset timing and persistence are more stable than a single terminal brightness value [19]. The shaded interval also shows why a late observation alone would compress the distinction among younger groups. The temporal encoder therefore represents onset timing, slope, and persistence separately; the empirical figure caption should additionally report the radiance unit and uncertainty-band definition.

Optical evidence resolves a different part of the ambiguity. Figure 3(b) shows median built-up indices of 0.68, 0.61, 0.58, 0.55, and 0.52 across the five age groups, with interquartile widths of 0.15, 0.12, 0.10, 0.12, and 0.16, respectively. The youngest group remains well above the 0.50 built-up threshold but has the widest upper spread, consistent with mixtures of bright roofs, exposed construction material, and incompletely paved parcels. The oldest group has a lower median and a long upper whisker, indicating that material weathering and dense redevelopment can produce both muted and renewed optical responses. The compact distributions for the 6-10 and 11-20-year groups make optical appearance useful for stabilized surfaces, but their overlap prevents a reliable year estimate by itself. Joint interpretation is therefore driven by complementarity at the sample level: temporal onset separates recent expansion, while optical texture and distribution shape constrain surfaces whose light histories have already saturated.

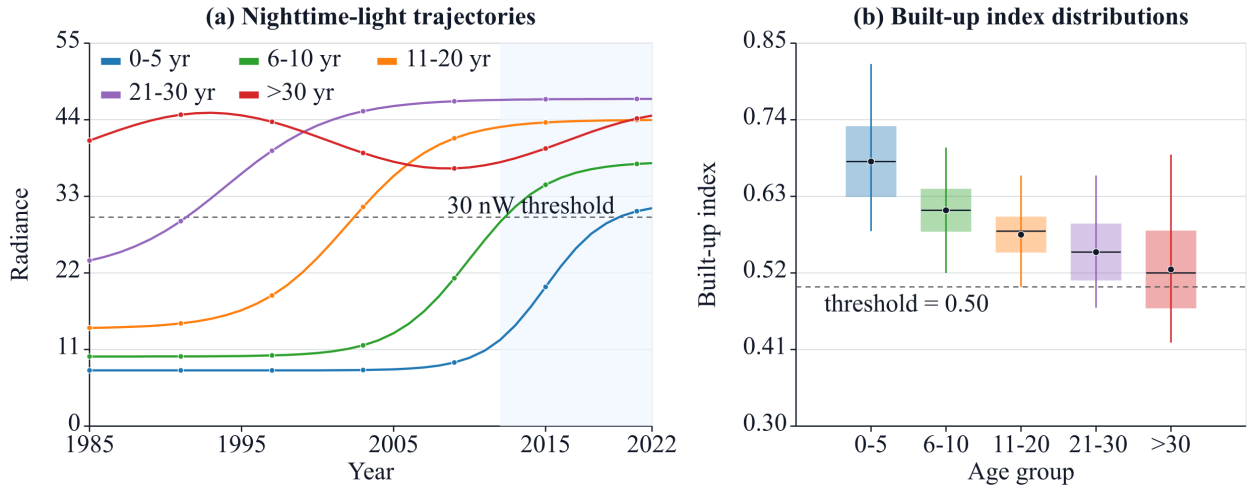


Figure 3. Temporal and optical separability: (a) annual nighttime-light trajectories by age group; (b) optical built-up index distributions by age group

The year 2022 will be designated as the target mapping year for the experiment. Therefore, the age range is from new-construction buildings to older than those observed in 1985. The three seasonal composites of the optical feature set are used to reduce phenological bias and avoid mistaking transient vegetation for stable urban development. Normalization of the night-light is performed before feature extraction, and all annual descriptors are calculated after outlier suppression. The resulting feature tensor has 38 temporal light values, 12 temporal derivatives and persistence features, 18 optical spectral and seasonal variables, 8 morphology variables and 4 neighborhood context variables. The last model has a small input for tile-based inference, but it still has a long history of signals.

Quality control excluded samples with continuous water occlusion, unstable cloud contamination, and irregular space mapping. Finally, the five age groups and the five morphological classes were kept. Newly built samples are deliberately oversampled in the mini-batch construction because they are less frequent and more likely to be mistaken for bare soil or temporary construction areas. The old samples are not oversampled because there are many in the Core and Inner

Ring Areas. Thus, rare-class learning can be achieved without ignoring the natural dominance of mature impervious surfaces in the benchmark.

The composition of the sample is also significantly older. The Core zone has 64.8% of its area older than 30 years, and the New Town has a younger-than-10-year-old-surface ratio of 37.5%. Industrial Belt has a high proportion of ambiguous confidence labels because the lighting in the industrial area is generally dim. Riverside is near many bridges and embankments, has shadows under the trees by the river, and has very reflective roofs. Airport Corridor has the highest base nighttime light intensity and is therefore more suitable for testing the model's detection of age from shape changes in light.

Temporal and optical distributions show complementary separability. Recent construction is characterized by a delayed, steep, and persistent rise in nighttime-light intensity, whereas surfaces older than 30 years provide little recent onset information because their radiance is already saturated. Optical built-up indices respond to material and seasonal variation: the youngest class is heterogeneous because bare soil, concrete, metal roofs, and partially paved surfaces can coexist, while mature classes have narrower but overlapping distributions. These patterns support joint inference, but they do not by themselves establish a physical construction date for redeveloped parcels.

5.2. Comparative Performance

Table 2 compares the proposed EfficientViT-BiFPN model with five baselines. The CNN baseline reaches 82.4% age-bin accuracy and a 4.8-year MAE, while U-Net reduces the MAE to 4.1 years. Swin-T and ConvNeXt reach 88.2% and 87.5% accuracy but require 34.7 and 31.5 ms per tile. The proposed model obtains 90.8% accuracy, 0.884 macro F1, a 2.7-year MAE, and a 4.3-year RMSE at 18.6 ms. Reporting categorical and continuous errors together prevents interval agreement from concealing large year-level deviations [20]. Relative to Swin-T, the proposed model gains 2.6 accuracy points while reducing latency by 46.4%. Hardware-consistent latency comparison is necessary when accuracy improvements are intended for repeated regional production [21].

Table 2. Comparative performance of age estimation models

Model	Accuracy (%)	Macro F1	MAE (yr)	RMSE (yr)	Latency (ms/tile)	Params (M)
Random Forest	76.9	0.731	5.6	8.2	11.4	2.1
CNN	82.4	0.795	4.8	7.0	16.8	8.7
U-Net	85.7	0.824	4.1	6.2	28.9	24.5
Swin-T	88.2	0.861	3.2	5.1	34.7	28.3
ConvNeXt-T	87.5	0.852	3.4	5.3	31.5	29.0
EfficientViT-BiFPN	90.8	0.884	2.7	4.3	18.6	17.6

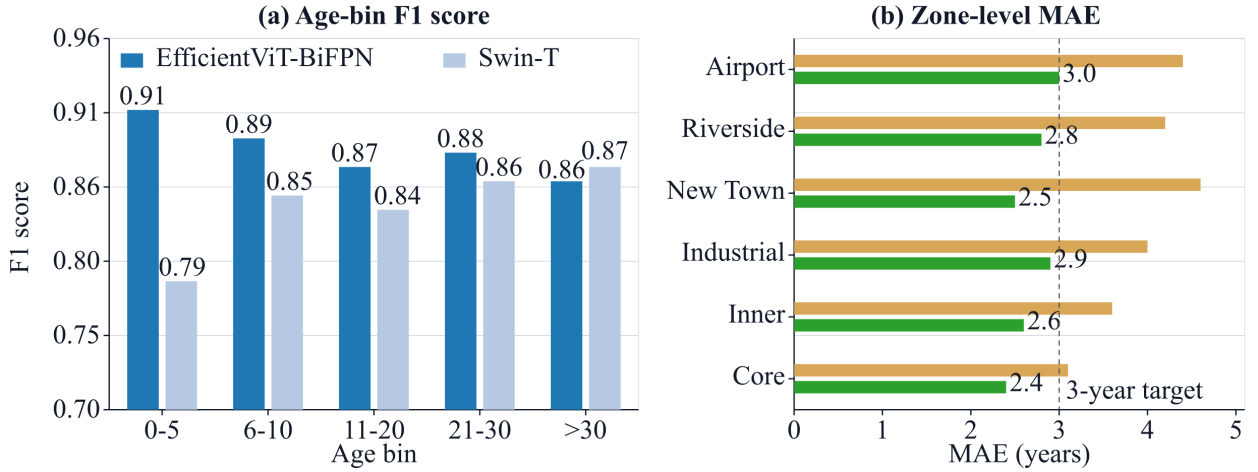


Figure 4. Comparative class and zone behavior: (a) F1 score by age bin; (b) MAE by urban zone

Figure 4(a) shows F1 scores of 0.91, 0.89, 0.87, 0.88, and 0.86 for the five age bins. The largest margin over Swin-T occurs for the 0–5-year bin, where F1 rises from 0.79 to 0.91; the margin contracts for older-than-30-year surfaces because both models benefit from persistent urban context. Gong et al. [22] demonstrated that class imbalance can substantially weaken rare-class segmentation, which makes class-wise reporting necessary when newly developed land is underrepresented. Pyramid feedback contributes most when optical edges are sharp but the nighttime-light trajectory

has not stabilized. Figure 4(b) shows that the model remains below the 3-year target in five zones and records 2.5 years in New Town, compared with 4.6 years for the strongest baseline. The smaller gain in Core is constrained by saturation and redevelopment, whereas the 2.1-year gain in New Town reflects a clearer progression from unlit land to constructed texture and sustained lighting.

Figure 5(a) shows diagonal accuracies of 91.7% for 0-5 years and 92.4% for surfaces older than 30 years, with the strongest exchange between the 11-20-year and 21-30-year bins. This pattern is consistent with boundary uncertainty, but directional bias should be evaluated with signed errors rather than inferred from visual symmetry. Figure 5(b) gives a fitted slope of 0.94 and a coefficient of determination of 0.86, indicating modest compression toward intermediate ages. Error diagnosis benefits from combining class confusion, continuous agreement, residual concentration, and spatial stratification [23]. Figure 5(c) places 68.5% of residuals in the central error interval and shows no secondary mode; the numerical bounds of that interval should be stated in the figure caption. Figure 5(d) records MAEs of 2.2 years for regular residential blocks and 3.6 years for fragmented commercial parcels, with uncertainty increasing from 2.2 to 4.4.

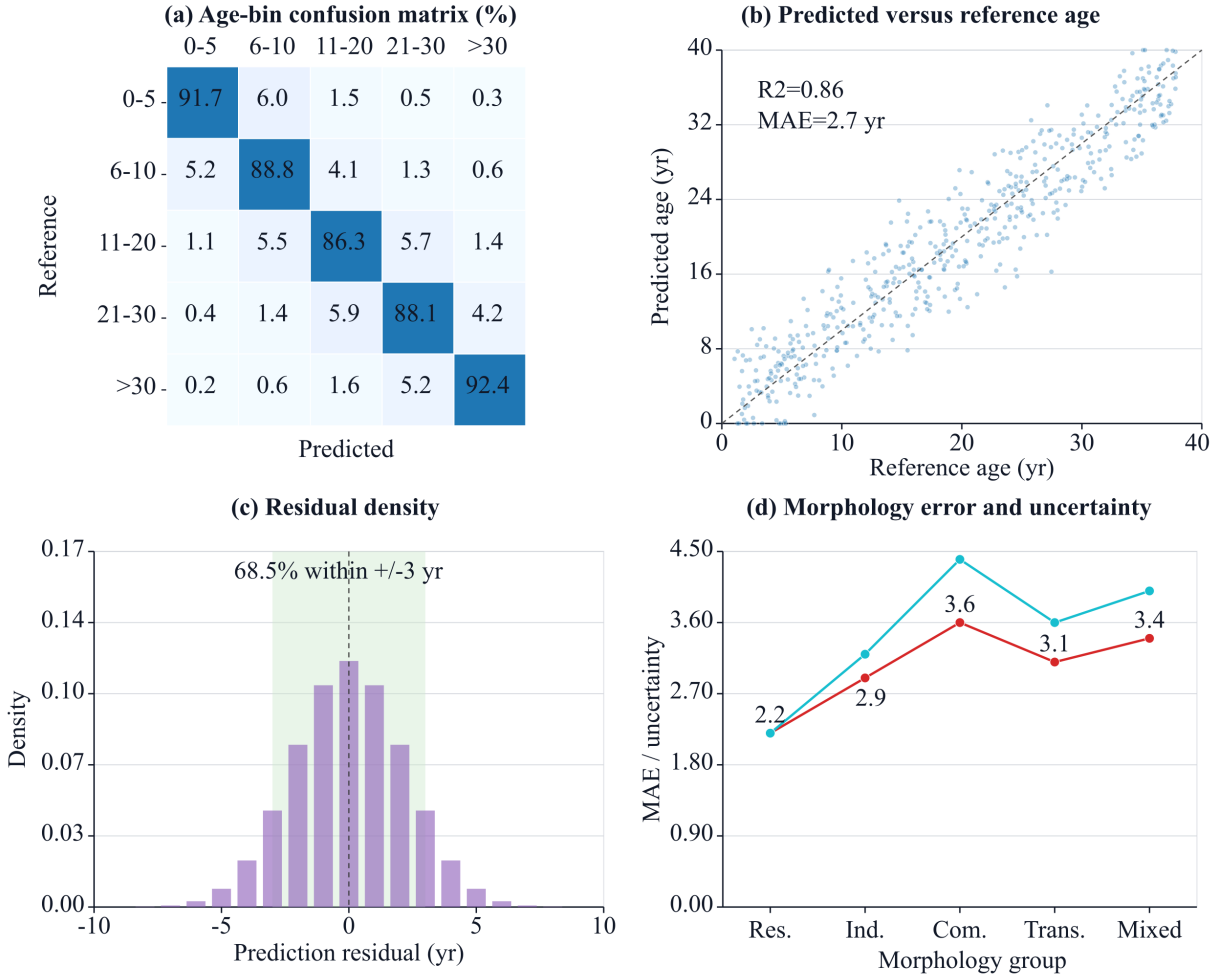


Figure 5. Error structure: (a) age-bin confusion matrix; (b) predicted versus reference age; (c) residual density; (d) morphology-group error

Random Forest provides a weak benchmark because engineered descriptors do not capture the same spatial representation available to image encoders. The CNN improves local texture recognition but has limited district-scale context. U-Net incorporates multi-scale convolutional features, although its top-down decoder does not explicitly retain repeated bidirectional interactions. Swin-T benefits from broader self-attention context but requires more memory, while ConvNeXt-T provides strong convolutional features with weaker transfer in the reserved morphology strata. These comparisons are descriptive point estimates; a definitive ranking requires identical inputs, augmentation, tuning budgets, hardware, and repeated-run uncertainty for every model.

The proposed model improves performance across recent, intermediate, and old age groups through different combinations of evidence. Optical texture and boundary-aware fusion are most useful for recent construction, temporal slope and neighborhood context support intermediate groups, and persistent light history helps distinguish older surfaces. The overall MAE of 2.7 years and RMSE of 4.3 years indicate that most predictions are close to the reference while a

smaller set of larger errors remains. Because only point estimates are available, these differences should be interpreted as benchmark trends rather than statistically established superiority.

The confusion matrix reveals that most errors occur between adjacent age bins, which is expected for an ordered target. The most consequential errors are cross-stage mistakes between young and old surfaces. These are relatively rare in the proposed model because the ordinal head discourages implausible jumps. The scatter and residual plots show mild underestimation for the oldest samples, reflecting the limited pre-1985 evidence. Morphology-group analysis shows that regular residential blocks are easiest because they combine stable texture and coherent light history. Fragmented commercial parcels are hardest because they undergo partial redevelopment and lighting changes that do not always coincide with surface construction.

5.3. Ablation Analysis and Model Behavior

Removing nighttime-light descriptors increases the MAE by 0.8 years and the late-stage confusion from 7.9% to 14.0%. Removing optical texture increases the young-surface error from 2.4 to 4.2 years, and removing BiFPN reduces the macro F1 from 0.884 to 0.846. One-factor-at-a-time ablation maintains an easy-to-understand view of the performance improvement due to an individual information source or fusion method [24]. A larger Transformer is not more accurate and is almost twice as slow. Light history helps to prevent the confusion of mature and old surfaces, and optical texture corrects problems in the early construction stage before radiance stabilizes.

Figure 6 reports the full model at 0.91 accuracy, 0.88 macro F1, 0.84 transfer stability, 0.86 calibration, 0.89 boundary consistency, and 0.82 efficiency-normalized utility. Removing nighttime light reduces transfer stability to 0.70; removing optical texture lowers boundary consistency to 0.72; and removing BiFPN reduces accuracy to 0.87 and boundary consistency to 0.74. Multi-criterion profiles are useful when components affect transfer, calibration, and boundaries differently from aggregate accuracy [25]. The full model does not maximize every normalized efficiency axis, but it provides the strongest quality-oriented balance within the tested latency budget; this interpretation remains conditional on the chosen axis definitions and normalization.

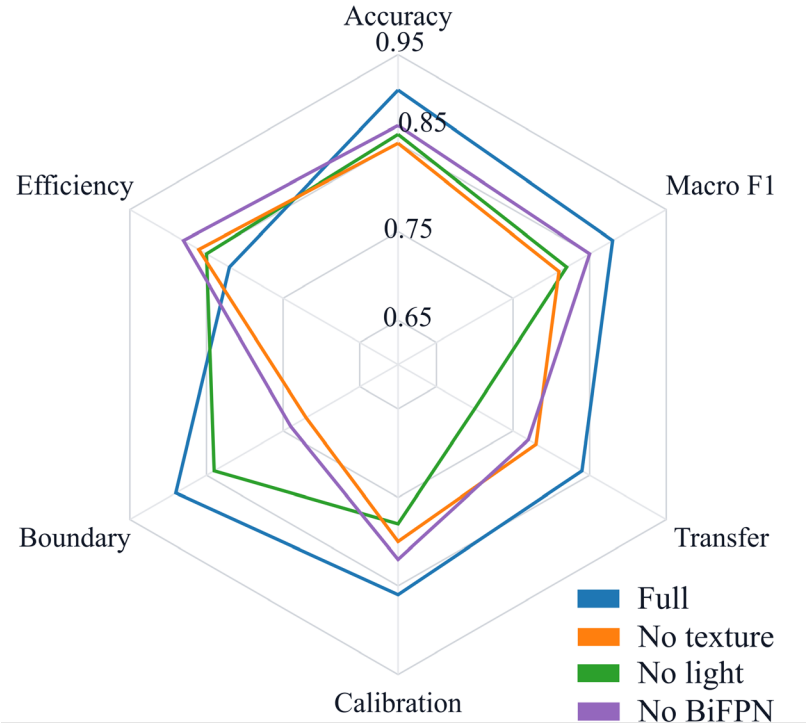


Figure 6. Ablation radar profile across accuracy, macro F1, transfer stability, calibration, boundary consistency, and efficiency-normalized utility

Figure 7(a) shows accuracies of 88.6% on Riverside and 87.9% on Airport Corridor, compared with 84.1% and 83.5% for Swin-T. Similar 4.4-4.5-point margins across two different zones indicate that the gain is not tied to one landscape. Figure 7(b) shows expected calibration errors of 0.041 and 0.047, both below the 0.05 target, while ConvNeXt reaches 0.073 and 0.078. Reserved-zone testing exposes geographic overfitting that random pixel partitions cannot reveal [26]. The slightly weaker Airport Corridor results identify elongated roads and intense linear lighting as harder conditions than compact riverside development. Confidence weighting preserves ambiguous samples without allowing them to dominate the decision surface.

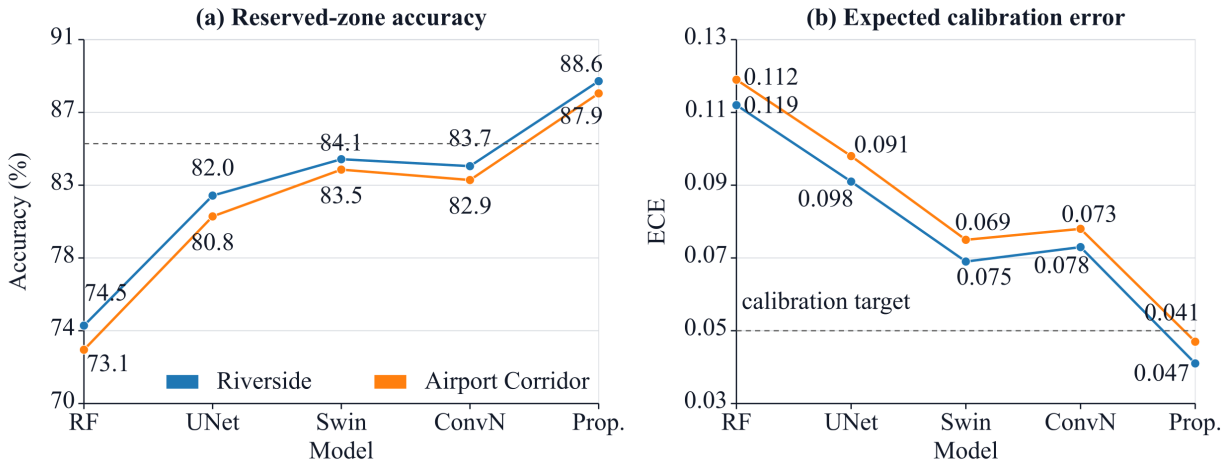


Figure 7. Cross-zone deployment: (a) reserved-zone accuracy comparison; (b) expected calibration error by reserved zone

Figure 8(a) shows accuracy increasing from 88.9% at 14.2 ms with one BiFPN repeat to 90.8% at 18.6 ms with three repeats, followed by only a 0.2-point gain at 22.9 ms. Figure 8(b) records a 1.42 GB peak memory footprint, 34.6% below Swin-T. Figure 8(c) gives 5.17 accuracy points per million parameters, above ConvNeXt and Swin-T but below the compact CNN, whose parameter advantage accompanies an 8.4-point accuracy deficit. Figure 8(d) places uncertainty at 0.22 for the 21–30-year group, above the 0.20 review threshold, while the youngest and oldest groups remain at 0.12 and 0.17. The three-repeat setting lies at the latency-accuracy knee, but 18-27-year redevelopment cases still require targeted review rather than acceptance based only on the point estimate.

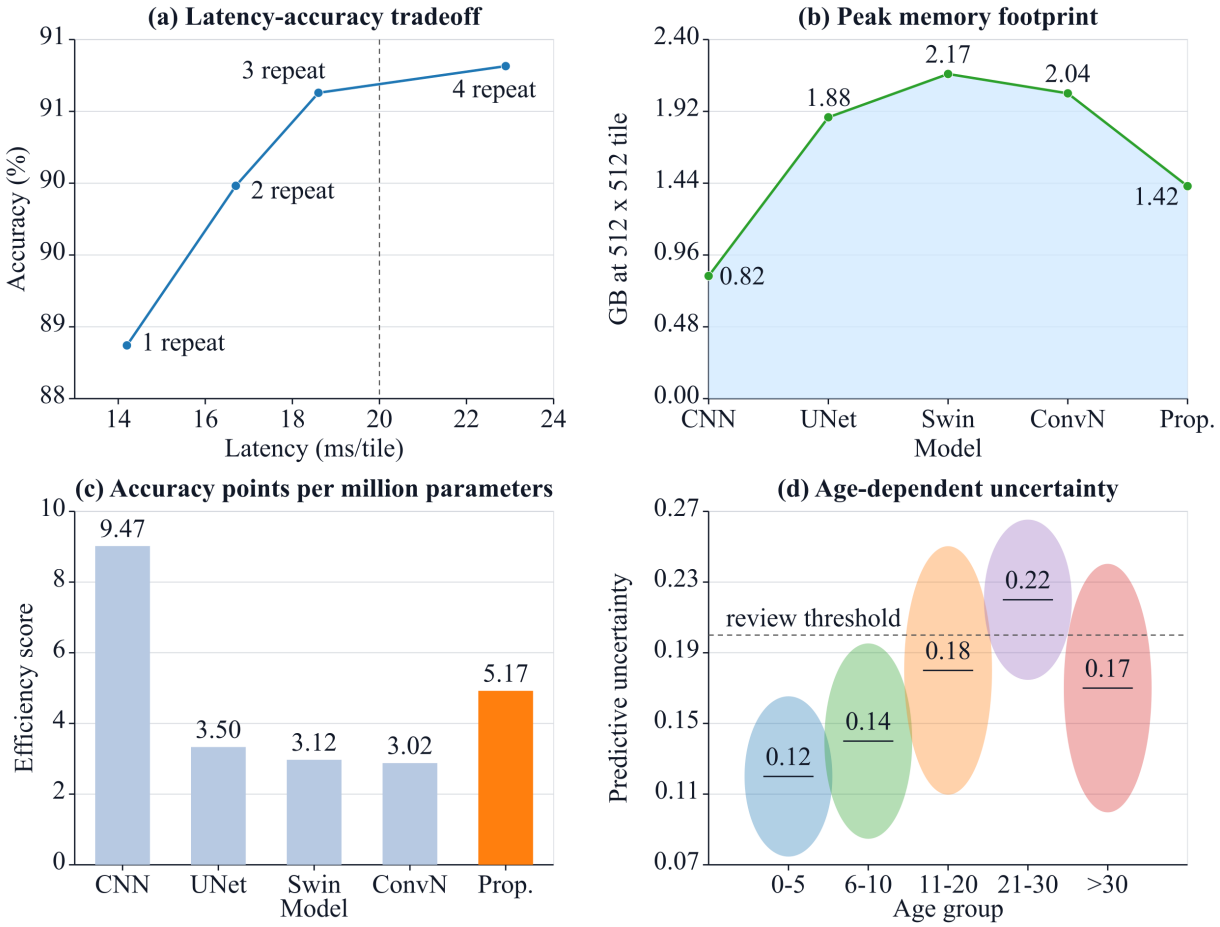


Figure 8. Computational and uncertainty behavior: (a) latency-accuracy tradeoff by BiFPN repeats; (b) peak memory by model; (c) parameter efficiency; (d) age-dependent uncertainty

The no-light ablation indicates that optical appearance alone has limited ability to distinguish surfaces whose materials have stabilized but whose chronology differs. Conversely, the no-texture variant retains broad development-stage information from illumination history but loses spatial specificity and material discrimination. Removing BiFPN

weakens multi-scale interaction and increases boundary error for linear infrastructure. These results demonstrate dependence on complementary inputs under the tested configuration, but they do not prove that the components are nonredundant in every dataset or architecture.

Uncertainty is highest for the 18-27-year interval, which contains many benchmark cases representing road widening, factory conversion, and commercial renovation. Recent surfaces can also be uncertain when construction remains incomplete and illumination slopes are irregular. Very old samples have an unknown pre-archive onset, although stable brightness and dense neighborhood context can still support coarse classification. The uncertainty layer should therefore accompany the age map and trigger targeted review, especially where redevelopment changes appearance without a corresponding illumination transition.

Efficiency analysis shows that the proposed configuration is suitable for repeated tile inference. Although the Random Forest baseline is faster, it lacks the spatial representation needed for accurate age mapping. The proposed model is substantially faster than Swin-T and ConvNeXt while maintaining higher accuracy. The three-repeat BiFPN setting is the best operating point because a fourth repeat adds 4.3 ms per tile for only 0.2 percentage points of accuracy. For a regional production run containing thousands of tiles and several preprocessing variants, this difference becomes important. The parameter-efficiency score further indicates that the accuracy gain is not simply purchased by model size.

6. Conclusion

This study developed and tested an efficientViT-BiFPN framework that integrates historical nighttime-light trajectory data with optical spectral-textural information to achieve accurate and efficient estimation of urban impervious surface age via compact multi-scale fusion.

1. The proposed model achieved 90.8% age-bin accuracy, a macro F1 score of 0.884, a mean absolute error of 2.7 years, and an RMSE of 4.3 years on 52,416 labeled samples from six urban zones. With 17.6 million parameters, it processed each 512 by 512 tile in 18.6 ms, placing the three-repeat BiFPN configuration at the practical knee of the latency-accuracy curve.

2. Relative to Swin-T, EfficientViT-BiFPN improved accuracy by 2.6 percentage points while reducing latency by 46.4%, and it retained 88.6% and 87.9% accuracy in the reserved Riverside and Airport Corridor zones. Removing nighttime-light descriptors increased MAE from 2.7 to 3.5 years and late-stage confusion from 7.9% to 14.0%, whereas removing BiFPN reduced macro F1 from 0.884 to 0.846. These contrasts show that illumination history separates mature surfaces, while optical texture and bidirectional pyramid fusion preserve recent-development cues and boundary detail.

3. The principal innovation is a confidence-aware coupling of lightweight local-global encoding with multi-scale temporal-optical fusion, with age treated as an ordered and continuous target rather than an extension of current impervious mapping. Its applicability depends on radiometric harmonization, persistent transition evidence, and optical observations that correspond to construction-related change. Redevelopment without a clear illumination or material transition remains difficult, and uncertainty for the 21-30-year group reaches 0.22, above the 0.20 review threshold. Sensor discontinuities, saturation, sparse cloud-free composites, and local illumination policies therefore require calibration and manual review. Because the benchmark is controlled and synthetic, these findings should not be interpreted as operational construction-date validation.

The framework establishes controlled benchmark evidence for interpretable urban age mapping. Future work should validate the method against independently dated building, road, and parcel archives; introduce explicit redevelopment labels and spatially buffered evaluation; report repeated-run uncertainty; and investigate domain adaptation and source-specific uncertainty modeling before regional operational deployment.

7. Declarations

7.1 Author Contributions

Galip Kahraman contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Data Availability

Not applicable.

7.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Ahmad, M. N., Shao, Z., Xiao, X., Fu, P., Javed, A., & Ara, I. (2024). A novel ensemble learning approach to extract urban impervious surface based on machine learning algorithms using SAR and optical data. International Journal of Applied Earth Observation and Geoinformation, 132, 104013. https://doi.org/10.1016/j.jag.2024.104013

[2] Javed, A., Shao, Z., Ara, I., Ahmad, M. N., Huq, Md. E., Saleem, N., & Karim, F. (2024). Development of Soil-Suppressed Impervious Surface Area Index for Automatic Urban Mapping. Photogrammetric Engineering & Remote Sensing, 90(1), 33–43. https://doi.org/10.14358/pers.23-00043r2

[3] Ren, H., Huang, X., Yang, J., & Zhou, G. (2025). Improving 30-meter global impervious surface area (GISA) mapping: New method and dataset. ISPRS Journal of Photogrammetry and Remote Sensing, 220, 354–376. https://doi.org/10.1016/j.isprsjprs.2024.12.023

[4] Woreket, W., & Zeleke, G. A. (2024). Extracting Urban Built-up Areas from Optical and Radar Data Fusion using Machine Learning Algorithms. International Journal of Image and Data Fusion, 15(2), 154–173. https://doi.org/10.1080/19479832.2024.2338736

[5] Shao, Z., Ahmad, M. N., & Javed, A. (2024). Comparison of Random Forest and XGBoost Classifiers Using Integrated Optical and SAR Features for Mapping Urban Impervious Surface. Remote Sensing, 16(4), 665. https://doi.org/10.3390/rs16040665

[6] Ling, J., Zhang, H., Liu, R., & Lin, Y. (2024). SCFN: A Deep Network for Functional Urban Impervious Surface Mapping Using C -Band and L -Band Polarimetric SAR Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 13149–13161. https://doi.org/10.1109/jstars.2024.3427839

[7] Marzi, D., Munoz Rios, E., Sorriso, A., Dell'Acqua, F., & Gamba, P. (2024). Joint multitemporal SAR and optical mapping of urban changes. International Journal of Microwave and Wireless Technologies, 16(3), 359–370. https://doi.org/10.1017/s175907872400031x

[8] Dermosinoglou, A., & Petropoulos, G. P. (2024). Exploring long term Impervious Surface Areas (ISA) dynamics using Landsat imagery, Machine Learning and GEE: The case of Attica, Greece. Remote Sensing Applications: Society and Environment, 36, 101338. https://doi.org/10.1016/j.rsase.2024.101338

[9] Zheng, Y., Cai, Y., Yang, K., Fan, M., Fu, M., & Wei, C. (2024). Quantitative spatiotemporal evolution of large urban agglomeration expansion based on 1995–2020 nighttime light and spectral data. Ecological Informatics, 84, 102824. https://doi.org/10.1016/j.ecoinf.2024.102824

[10] Yu, T., Liu, C., Li, W., Huang, W., Wu, H., & Fan, Z. (2024). Characterizing urban actively populated area growth in the Yangtze River Delta using nighttime light data. International Journal of Applied Earth Observation and Geoinformation, 129, 103857. https://doi.org/10.1016/j.jag.2024.103857

[11] Shi, K., Pan, Y., Jiang, L., Wang, J., Cui, Y., Dong, T., Ma, J., & Huang, C. (2024). A Nighttime Light Based Urban Sprawl Model Revealing Reduced Electricity Intensity With Increasing Urban Size. IEEE Geoscience and Remote Sensing Letters, 21, 1–5. https://doi.org/10.1109/lgrs.2024.3352644

[12] Dong, B., Zhang, R., Li, S., Ye, Y., & Huang, C. (2025). A meta-analysis for the nighttime light remote sensing data applied in urban research: Key topics, hotspot study areas and new trends. Science of Remote Sensing, 11, 100186. https://doi.org/10.1016/j.srs.2024.100186

[13] Adegun, A., Viriri, S., & Tapamo, J.-R. (2024). Automated classification of remote sensing satellite images using deep learning based vision transformer. Applied Intelligence, 54(24), 13018–13037. https://doi.org/10.1007/s10489-024-05818-y

[14] Papa, L., Russo, P., Amerini, I., & Zhou, L. (2024). A Survey on Efficient Vision Transformers: Algorithms, Techniques, and Performance Benchmarking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 7682–7700. https://doi.org/10.1109/tpami.2024.3392941

[15] Luo, W., Deng, F., Jiang, P., Dong, X., & Zhang, G. (2024). FSegNet: A Semantic Segmentation Network for High-Resolution Remote Sensing Images That Balances Efficiency and Performance. IEEE Geoscience and Remote Sensing Letters, 21, 1–5. https://doi.org/10.1109/lgrs.2024.3398804

[16] Nedevschi, S. (2024). Semantic segmentation of remote sensing images with transformer-based U-Net and guided focal-axial attention. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 18303-18318. https://doi.org/10.1109/jstars.2024.3470316

[17] Boulila, W., Ghandorh, H., Masood, S., Alzahem, A., Koubaa, A., Ahmed, F., Khan, Z., & Ahmad, J. (2024). A transformer-based approach empowered by a self-attention technique for semantic segmentation in remote sensing. Heliyon, 10(8), e29396. https://doi.org/10.1016/j.heliyon.2024.e29396

[18] Zhao, F., Zhu, Z., Liu, H., & Zhang, J. (2024). Multiscale Alignment and Progressive Feature Fusion Network for High Resolution Remote Sensing Images Change Detection. IEEE Transactions on Geoscience and Remote Sensing, 62, 1–13. https://doi.org/10.1109/tgrs.2024.3433373

[19] Guo, Z., Chen, H., & He, F. (2025). MSFNet: Multiscale Spatial-Frequency Feature Fusion Network for Remote Sensing Change Detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 1912–1925. https://doi.org/10.1109/jstars.2024.3481424

[20] Liu, P. (2025). A multimodal fusion framework for semantic segmentation of remote sensing based on multilevel feature fusion learning. Neurocomputing, 653, 131233. https://doi.org/10.1016/j.neucom.2025.131233

[21] Zermatten, V., Lu, X., Castillo-Navarro, J., Kellenberger, T., & Tuia, D. (2024). Land Cover Mapping From Multiple Complementary Experts Under Heavy Class Imbalance. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 6468–6477. https://doi.org/10.1109/jstars.2024.3369876

[22] Gong, Z., Duan, L., Xiao, F., & Wang, Y. (2024). MSAug: Multi-Strategy Augmentation for rare classes in semantic segmentation of remote sensing images. Displays, 84, 102779. https://doi.org/10.1016/j.displa.2024.102779

[23] Ferchichi, A., Ferchichi, A., Hendaoui, F., Chihaoui, M., & Toujani, R. (2025). Deep Learning-based Uncertainty Quantification for spatio-temporal environmental Remote Sensing: A systematic literature review. Neurocomputing, 639, 130242. https://doi.org/10.1016/j.neucom.2025.130242

[24] Zheng, J., Wen, Y., Chen, M., Yuan, S., Li, W., Zhao, Y., Wu, W., Zhang, L., Dong, R., & Fu, H. (2024). Open-set domain adaptation for scene classification using multi-adversarial learning. ISPRS Journal of Photogrammetry and Remote Sensing, 208, 245–260. https://doi.org/10.1016/j.isprsjprs.2024.01.015

[25] Ismael, S. F., Kayabol, K., & Aptoula, E. (2024). Unsupervised domain adaptation for the semantic segmentation of remote sensing images via a class-aware Fourier transform and a fine-grained discriminator. Digital Signal Processing, 151, 104551. https://doi.org/10.1016/j.dsp.2024.104551

[26] Silva Filho, P., Persello, C., Maretto, R. V., & Machado, R. (2024). Mapping the Brazilian savanna's natural vegetation: A SAR-optical uncertainty-aware deep learning approach. ISPRS Journal of Photogrammetry and Remote Sensing, 218, 405–421. https://doi.org/10.1016/j.isprsjprs.2024.09.019