

Review

Not peer-reviewed version

Machine Learning and Deep Learning Approaches in Thermal Remote Sensing: A Systematic Review (2018–2026)

[Azad Rasul](#) *

Posted Date: 31 March 2026

doi: 10.20944/preprints202603.2395.v1

Keywords: thermal remote sensing; machine learning; deep learning; land surface temperature; PRISMA 2020; convolutional neural network



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Review

Machine Learning and Deep Learning Approaches in Thermal Remote Sensing: A Systematic Review (2018–2026)

Azad Rasul ^{1,2}

¹ Department of Geography, Soran University, Soran, Erbil, Iraq

² Department of Forestry, College of Agricultural Engineering Sciences, Salahaddin University-Erbil, 44002, Iraq

* Correspondence: azad.rasul@soran.edu.iq

Abstract

Background: The intersection of machine learning (ML) and deep learning (DL) with thermal remote sensing (TRS) has undergone a transformative expansion since 2018, driven by the proliferation of high-resolution satellite missions and open-source deep learning frameworks. Despite this rapid growth, to the best of our knowledge, no comprehensive PRISMA-compliant systematic review has synthesised ML/DL applications specifically within the thermal RS domain across the post-2018 period. **Objectives:** This review maps the complete landscape of ML/DL applications in thermal RS from January 2018 to March 2026 with five primary objectives: (i) quantify publication trends; (ii) classify the taxonomy of ML/DL architectures; (iii) map application domain coverage; (iv) appraise methodological quality and open science practices; and (v) identify research gaps and future directions. **Methods:** A systematic electronic search was conducted across Scopus and Google Scholar. Following PRISMA 2020 guidelines (Page et al., 2021), records underwent a structured multi-stage screening process implemented in Python. This consisted of five main screening stages after initial deduplication, followed by a final full-text eligibility assessment for open-access records retrieved via the Unpaywall API. Data extraction employed a structured template covering bibliographic metadata, sensor platforms, ML/DL architecture, application domain, performance metrics, and open science practices. **Results:** A total of **193 peer-reviewed studies** met the inclusion criteria, of which **93 were available as open-access full texts** and **100 were accessible through title, abstract, and structured metadata only** due to institutional access restrictions. CNNs (43.7%), LSTM/BiLSTM (33.0%), and SVR/SVM (29.1%) were the dominant architectures across the 93 open-access full-text studies from which comprehensive architecture data were extracted. Application domains concentrated on SST forecasting, LST retrieval, LST downscaling, and gap-filling, leaving wildfire detection, evapotranspiration estimation, and permafrost monitoring relatively underrepresented compared to core domains. Code availability was reported in fewer than 5% of included studies. **Conclusions:** This review reveals a maturing but architecturally conservative field with transformative opportunities in physics-informed neural networks, transformer-based models, and underserved application domains. The persistent open science deficit represents a structural reproducibility challenge that warrants urgent community attention.

Keywords: thermal remote sensing; machine learning; deep learning; land surface temperature; PRISMA 2020; convolutional neural network

1. Introduction

1.1. Thermal Remote Sensing: Foundations and Significance

Thermal remote sensing (TRS) is a branch of Earth observation that measures emitted thermal infrared (TIR) radiation in the 8–14 μm atmospheric window to derive fundamental surface thermodynamic properties, most notably land surface temperature (LST) and sea surface temperature (SST). LST regulates sensible and latent heat flux exchanges between the land surface and the atmosphere, exerts a first-order control on evapotranspiration, and serves as a critical boundary condition for land surface and climate models (Li et al., 2013; Norman and Becker, 1995). SST plays a central role in air–sea heat exchange, ocean circulation, and the modulation of global weather and climate variability (Reynolds et al., 2002).

The retrieval of LST and emissivity from satellite-borne TIR sensors has historically relied on physics-based algorithms. Among the most influential is the generalised split-window algorithm of Wan and Dozier (1996), which underpins the standard MODIS LST product (MOD11). For multi-band TIR systems such as ASTER, the temperature–emissivity separation (TES) algorithm of Gillespie et al. (1998) provides simultaneous retrieval of surface temperature and emissivity spectra across five TIR bands. While physics-based algorithms have delivered reliable retrievals under well-characterised atmospheric and surface conditions, they are inherently sensitive to assumptions about atmospheric profiles and emissivity priors, and their performance degrades significantly under cloudy skies, complex terrain, and heterogeneous surface cover.

The landscape of satellite TRS has evolved substantially since 2018. The launch of ECOSTRESS aboard the International Space Station in June 2018 provided 70-m resolution TIR observations with irregular diurnal sampling — a capability not previously available from polar-orbiting missions (Fisher et al., 2020). Concurrently, Landsat 8 and 9 TIRS have continued to deliver 100-m thermal data with a 16-day revisit cycle (Xu et al., 2024), while Sentinel-3 SLSTR and GOES-R ABI extend coverage to finer temporal resolutions. This expanding constellation has generated unprecedented volumes of TIR data, demanding more sophisticated analytical frameworks than traditional physics-based methods can provide (He et al., 2025; Chang et al., 2024).

1.2. The Rise of Machine Learning and Deep Learning in Remote Sensing

The emergence of deep learning as a transformative paradigm was catalysed by the landmark review of LeCun et al. (2015) in *Nature*. The proliferation of open-source frameworks (TensorFlow, PyTorch) from 2015 onwards democratised access to these methods. In the Earth observation community, Zhu et al. (2017) established that deep learning architectures — CNNs, RNNs, and autoencoders — were achieving performance gains over traditional methods for tasks as diverse as scene classification, hyperspectral analysis, and change detection.

Since 2018, ML/DL adoption in thermal RS has expanded across all major application domains. In LST retrieval, Ye et al. (2025) demonstrated end-to-end CNN retrieval directly from satellite radiances, while Jia et al. (2021) coupled CNN and LSTM for spatiotemporal LST prediction over Beijing. In SST forecasting, Fu et al. (2024) applied deep learning to coastal temperature prediction in the South China Sea, and Song et al. (2024) introduced the STVformer transformer for multivariate SST forecasting. However, this growth has been uneven: wildfire detection, evapotranspiration estimation at field scale, and permafrost temperature monitoring remain substantially underserved.

1.3. Research Gap and Motivation

Despite this rapid proliferation, no comprehensive systematic review has synthesised ML/DL applications specifically within the thermal RS domain across the post-2018 period. Existing reviews address ML in broader remote sensing contexts without thermal specificity (Zhu et al., 2017), or focus narrowly on single applications such as urban heat islands (Zhou et al., 2019; Rasul et al., 2017). There

is therefore a clear and timely need for a rigorous, PRISMA-compliant systematic review that maps the entire landscape of ML/DL in thermal RS from 2018 to early 2026.

1.4. Objectives

This systematic review has five primary objectives:

- To identify and quantify the volume, geographic distribution, and temporal trends of publications applying ML/DL in thermal RS from 2018 to March 2026.
- To classify and map the full taxonomy of ML/DL architectures applied in TRS, including supervised, unsupervised, semi-supervised, and deep learning methods.
- To evaluate the range of thermal RS applications addressed and assess which are underserved by ML/DL approaches.
- To critically appraise methodological quality, reproducibility, and open-data/code practices.
- To identify research gaps, emerging trends, and priority directions for future ML/DL applications in thermal remote sensing.

2. Methodology

This review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021). The complete search, deduplication, and screening workflow was implemented in Python using Jupyter notebooks, ensuring full reproducibility and a transparent audit trail of all decisions.

2.1. Search Strategy and Information Sources

A systematic electronic search was conducted across two bibliographic databases: Scopus and Google Scholar. The search was executed in March 2026, covering publications from January 2018 to March 2026. The 2018 lower boundary was selected to coincide with the operational maturity of the Landsat 8 TIRS, the launch of ECOSTRESS, and the widespread adoption of deep learning frameworks in geospatial applications.

Scopus records were retrieved programmatically via the Elsevier Scopus API using paginated HTTP requests. Google Scholar records were retrieved using the scholarly Python library with inter-request delays of 8–15 seconds (randomised) to comply with rate-limiting policies.

2.1.1. Machine Learning / Deep Learning Keyword Block

The ML/DL block targeted a comprehensive range of algorithmic families. Keywords included: "machine learning", "deep learning", "neural network", "convolutional neural network", "CNN", "LSTM", "random forest", "support vector", "XGBoost", "transformer", "U-Net", "vision transformer", "GAN", "autoencoder", "physics-informed neural network", and "PINN".

2.1.2. Thermal Remote Sensing Keyword Block

The thermal RS block was designed to capture studies across the full spectrum of sensors, derived products, and application domains. Key terms included: "thermal remote sensing", "land surface temperature", "LST", "thermal infrared", "TIR", "heat island", "ECOSTRESS", "sea surface temperature", "SST", "Landsat TIRS", "MODIS LST", "Sentinel SLSTR", "ASTER TIR", "brightness temperature", and "thermal downscaling".

2.2. Eligibility Criteria

2.2.1. Inclusion Criteria

A study was eligible for inclusion if it satisfied all four criteria simultaneously:

- Application of at least one ML or DL method as a core analytical component — not merely referenced in passing.
- Use of thermal remote sensing data as primary input — including satellite-derived LST, SST, TIR imagery, brightness temperature, or data from thermal-band sensors.
- Publication between January 2018 and March 2026.
- Primary research articles or systematic reviews published in peer-reviewed journals, conference proceedings.

2.2.2. Exclusion Criteria

A study was excluded if any one of the following conditions applied:

- No ML/DL method employed — reliance solely on physics-based retrieval algorithms or classical statistical models.
- Thermal data absent — studies using only visible, multispectral (non-TIR), SAR, or LiDAR data.
- Publication year outside 2018–2026.
- Not a scientific publication — editorials, opinion pieces, or non-peer-reviewed grey literature.
- Non-Earth-observation thermal applications — medical thermography, industrial inspection, building energy simulation, or indoor thermal comfort.
- Thermal data derived exclusively from ground-based sensors, UAV/drone-mounted cameras, or handheld infrared cameras without satellite integration.
- Ocean numerical model outputs (e.g., FVCOM, ROMS, HYCOM) without satellite thermal RS data as input.

2.3. Study Selection Process

Study selection followed a structured, multi-stage screening process implemented in Python using Jupyter notebooks. This process consisted of five main screening stages (detailed below) after initial deduplication, followed by a final full-text eligibility assessment. The workflow was designed to ensure transparency, reproducibility, and rigorous application of the PRISMA 2020 guidelines (Page et al., 2021). The detailed record flow across all steps is presented in Table 1.

2.3.1. Stage 1 – Deduplication

Records from all databases were merged into a single master dataframe. Deduplication was performed via exact DOI matching followed by fuzzy title matching using the fuzzywuzzy Python library (Levenshtein distance ratio threshold of 88%), ensuring cross-database duplicates were removed regardless of DOI metadata availability.

2.3.2. Stages 2–4 – Time Filter, Title Screening, and Abstract Screening

Following deduplication, all records outside the 2018–2026 window were excluded. Each remaining record was assessed at title level for co-occurrence of ML/DL and thermal RS keywords. Records containing only one keyword category were forwarded as “uncertain” for abstract review. Abstract screening applied a tightened second-pass requiring confirmation of a specific ML method and a specific thermal sensor or product, with hard exclusion signals applied to the combined text.

2.3.3. Stage 5 – Domain Filter and Relevance Scoring

Surviving records underwent a domain exclusion filter targeting false-positive categories (energy systems, medical applications, engineering/industrial settings, non-satellite platforms). A relevance score was computed as a weighted sum across three keyword categories: high-value ML/DL terms (+3 each), high-value thermal RS terms (+4 each), and core application terms (+2 each),

with a title-level bonus of +2 and a weak co-occurring domain penalty of -1. Only records in the CORE (≥ 15 points) and HIGH (10–14 points) tiers were forwarded for full-text review.

2.4. Full-Text Screening and Data Access

Open-access PDFs were retrieved automatically using the **Unpaywall API** (api.unpaywall.org/v2/{doi}). Downloaded files were validated against the PDF magic bytes signature (%PDF). Full-text extraction used the *pdfplumber* Python library to extract Abstract and Methods sections from the first six pages of each PDF.

Data Access Limitation and Transparency Statement: Of the 193 studies meeting the inclusion criteria, **93 (48.2%) were available as open-access full texts**, from which comprehensive data extraction was performed across all fields. The remaining **100 studies (51.8%) were not accessible as full text** due to institutional paywalls at the time of retrieval. For these studies, data extraction was limited to title, author-provided abstract, and structured metadata fields obtainable via the Scopus API. Critically, for 84 of these 100 non-open-access (non-OA) studies, the ML/DL method and application domain were identifiable from title and abstract metadata; these studies are therefore included in descriptive analyses of publication trends, architecture classification, and application domain mapping. They are **excluded** from quantitative performance benchmarking and detailed methodological quality assessment where full-text data are required. This approach is consistent with PRISMA 2020 reporting practice, where studies with insufficient data for quantitative synthesis are retained in descriptive analyses (Page et al., 2021).

2.5. Data Extraction Framework

Data were extracted from all included studies using a structured template developed prior to screening. Fields captured included:

- Bibliographic information: title, authors, year, journal/conference, DOI, source database.
- Sensor and data characteristics: satellite/sensor platform, thermal band or product, spatial and temporal resolution.
- ML/DL method: primary architecture, deep learning framework, and training approach.
- Application domain: primary application (LST retrieval, LST downscaling, UHI mapping, SST forecasting, evapotranspiration, glacier monitoring, etc.).
- Performance metrics: RMSE, MAE, R^2 , and domain-specific metrics.
- Open science practices: availability of code and data, repository URLs.

2.6. PRISMA 2020 Record Flow

Table 1. PRISMA 2020 Record Flow – ML & Deep Learning in Thermal Remote Sensing (2018–2026).

PRISMA Stage		Records (n)	Notes
Phase 1 – Identification			
Records identified	(all databases)	5,407	Scopus API + Google Scholar (scholarly Python library)
Phase 2 – Deduplication			
Records after deduplication		5,386	21 duplicates removed: DOI exact match → fuzzy title match (Levenshtein threshold 88%)

PRISMA Stage	Records (n)	Notes
Phase 3 – Automated Keyword Screening		
After title screening	3,407	Retained if title contained ≥ 1 ML/DL keyword AND ≥ 1 thermal RS keyword; remainder excluded
After abstract screening	469	Required: specific ML method confirmed + specific thermal sensor/product; hard exclusion signals applied
Phase 4 – Domain Filter & Relevance Scoring		
After domain & relevance pre-screening	361	Non-EO, engineering, medical, and industrial domains removed; satellite RS platform confirmed
After bulk exclusion (UNCERTAIN / MEDIUM / LOW tiers)	250	Lower-priority relevance tiers removed; only CORE + HIGH tiers forwarded to full-text review
Phase 5 – Full-Text Eligibility Assessment		
Excluded at full-text eligibility assessment	-57	<i>Book chapters / non-peer-reviewed (n=7); non-English (n=7); no satellite thermal IR data (n=18); retracted articles (n=2); unverifiable / truncated metadata from API errors (n=23)</i>
Phase 6 – Included Studies		
Studies included – with full PDF (open access)	93	OA PDFs retrieved via Unpaywall API; abstract, methods, and performance metrics extracted
Studies included – metadata only (non-OA)	100	Included via title/abstract and DOI metadata only; excluded from quantitative performance benchmarking
FINAL INCLUDED TOTAL	193	All 193 studies included in synthesis; stratified by data availability for performance analysis

2.7. Quality Assessment and Risk of Bias

Methodological quality of included studies was assessed using a domain-adapted checklist informed by the Cochrane Risk of Bias framework and reproducibility standards for ML-based geospatial research. Each study was evaluated on: (i) clarity and completeness of ML/DL architecture description; (ii) independence of training and test datasets; (iii) appropriateness of performance metrics; (iv) generalisability of validation (spatial and temporal cross-validation); (v) availability of code and data; (vi) comparison against a baseline or benchmark method; and (vii) transparency of sensor pre-processing and atmospheric correction procedures.

2.8. Synthesis Approach

Given the substantial methodological heterogeneity across ML architectures, sensor platforms, and application domains, a semi-quantitative synthesis approach was adopted rather than meta-analysis. Quantitative data extraction (RMSE, MAE, R^2) was performed to allow structured comparison within application subgroups, but pooled statistical synthesis was not attempted given the incompatibility of performance metrics across tasks. Synthesis is structured around five analytical axes: (i) temporal publication trends, (ii) ML/DL architecture taxonomy, (iii) application domain coverage, (iv) sensor and platform usage, and (v) open science practices.

3. Results

A total of **193 peer-reviewed studies** published between 2018 and 2026 were included in this systematic review following the full PRISMA 2020 screening pipeline. Of these, **93 studies were available as open-access full texts** and **100 studies were included on the basis of title, abstract, and structured metadata** due to institutional access limitations at the time of retrieval. The 193 included studies collectively span four principal thermal RS domains: LST retrieval, LST downscaling and spatiotemporal fusion, LST and SST gap-filling and reconstruction, and SST forecasting, with additional coverage of urban heat island (UHI) dynamics, agricultural and drought monitoring, and permafrost/glacier thermal studies.

3.1. Publication Trends (2018–2026)

The temporal distribution of included studies reveals a clear and sustained exponential growth trajectory in the application of ML and DL to thermal remote sensing over the review period (Table 2). Publication output was modest in the early years, with only 1 study in 2018 and 5 in 2019, reflecting the nascent adoption of deep learning in the thermal RS community (Shatnawi and Abu Qdais, 2019; Choe and Yom, 2020). Growth remained gradual through 2022, with 5 studies in 2020, 4 in 2021, and 7 in 2022, bringing the cumulative total to 22 by end of that year.

A pronounced and sustained surge began from 2023 onwards: 25 studies in 2023, 62 in 2024, 65 in 2025, and 19 studies already published in the first months of 2026 — yielding a cumulative total of 193 included studies across the review period.

This acceleration is consistent with three converging drivers: the availability of large open-access thermal satellite archives (ECOSTRESS, Landsat Collection 2, MODIS Version 6.1) (Xu et al., 2024; Chang et al., 2024); the maturation of PyTorch and TensorFlow ecosystems for spatiotemporal modelling; and the introduction of vision transformers and their domain-adapted variants into remote sensing workflows (Hu et al., 2025; Song et al., 2024; Guo et al., 2025).

Table 2. Annual publication counts of included studies applying ML/DL in thermal remote sensing (2018–2026). Year metadata was available for all 193 included studies.

Year	Studies (n)	Cumulative (n)	Notable Trend
2018	1	1	Early adoption — LSTM and CNN pioneers
2019	5	6	SVR and RF baselines; first ANN-LST studies
2020	5	11	CNN+LSTM spatiotemporal coupling
2021	4	15	Continued growth; ensemble and hybrid methods
2022	7	22	3D-CNN and spatiotemporal fusion growth; XGBoost-LST
2023	25	47	Transformer models enter thermal RS; systematic literature reviews
2024	62	109	ViT, GAN super-resolution, ECOSTRESS uptake; graph neural networks
2025	65	174	Foundation models, physics-informed NNs, U-Net variants
2026	19	193	Hierarchical transformers, SST forecasting, XGBoost-SST retrieval

3.2. Taxonomy of ML/DL Architectures

Architecture data were extractable for **177 of 193 included studies** (93 full-text records + 84 non-OA records where the ML method was identifiable from title/abstract metadata). A rich and diverse ecosystem of ML and DL architectures was identified, classified into five broad families. Multiple architectures were frequently combined within single studies, reflecting the trend towards hybrid and ensemble modelling (Fu et al., 2024; Zhang et al., 2025; Rani et al., 2025).

3.2.1. Deep Convolutional Architectures

Convolutional neural networks (CNNs) constituted the single most frequently employed architecture, appearing in 45 of the 93 full-text studies (43.7%). Their dominance reflects the strong spatial structure of thermal RS data, where local pixel-neighbourhood relationships in LST and SST fields are efficiently captured by convolutional filters (Tanoori et al., 2024). Standard 2D CNNs were applied across all application domains, while 3D-CNN variants were particularly prevalent in SST reconstruction tasks where temporal stack inputs from geostationary sensors (GOES-R, Himawari AHI) were used.

Encoder-decoder architectures, principally U-Net and its variants (U-Transformer, ST-UNet, the novel U-MoE adaptive mixture-of-experts U-Net of Zhang et al., 2026), appeared in 7 full-text studies and were strongly favoured for spatially dense prediction tasks — LST downscaling, SST gap-filling, and mesoscale eddy detection. Sun and Wang (2025) introduced a coordination attention residual U-Net for SST reconstruction in the South China Sea, while Qi et al. (2025) proposed the

RVDCU-Net framework for spatiotemporal SST reconstruction combining residual and dilated convolution modules.

Generative adversarial networks (GANs) appeared in 11 full-text studies (10.7%). They were mainly used for thermal super-resolution and SST downscaling. Wang et al. (2024) used CycleGAN to correct climate model SST simulations. It improved climatological biases, interannual variability including ENSO, and SST extremes. Jalbuena et al. (2025) applied a Regression-based GAN (RGAN) to fix landcover bias in MODIS LST downscaling. It outperformed traditional methods across urban areas in Korean cities. Sundar and Patchaiammal (2024) developed a hybrid GAN-RNN model. It integrated remote sensing data for urban environmental analysis and resilience. GANs are one of the fastest-growing methods. They had no uses before 2022 and nearly doubled in publications during 2023–2024.

Autoencoder architectures (including DINCAE and its Inception-enhanced variant I-DINCAE) appeared in 4 full-text studies, exclusively in SST and LST gap-filling contexts, proving well-suited to cloud-gap reconstruction where the objective is to infer missing thermal values from partial spatial observations (Barth et al., 2020).

3.2.2. Recurrent and Sequence Models

Long short-term memory (LSTM) networks and their bidirectional variants (BiLSTM) were the dominant recurrent architecture, appearing in 34 full-text studies (33.0%), of which 7 employed ConvLSTM. LSTM models were most prevalent in SST forecasting tasks, where temporal autocorrelation at daily-to-monthly timescales was exploited (Cai et al., 2025). Li et al. (2026) used LSTM to predict urban LST in the context of land use change scenarios, while Yin et al. (2025) proposed AIGR-LSTM for LST interpolation combining adaptive information granulation with recurrent learning. El Azhary and Minaoui (2025) proposed an encoder-decoder dual attention ConvLSTM architecture for **Moroccan coastal SST forecasting**, combining convolutional operations for spatial dependencies and LSTM for temporal sequences with a dual attention mechanism (contextual + spatial), achieving significant improvements in prediction accuracy and computational efficiency compared to single-attention baselines.

3.2.3. Transformer-Based Architectures

Transformer architectures emerged as a rapidly growing architectural family, appearing in 13 full-text studies. Their adoption followed a clear chronological trajectory: absent before 2022, first appearing in SST forecasting applications in 2023, and expanding to LST retrieval and downscaling by 2024–2026. Jia et al. (2024) introduced TL-iTransformer for SST prediction with transfer learning, while Song et al. (2024) proposed the STVformer spatial-temporal-variable transformer for improved SST forecasting. Hu et al. (2025) proposed THSTNet, a two-stage hierarchical spatiotemporal fusion network using the Swin Transformer architecture to fuse MODIS and Landsat LST products for fine-resolution (30-m) LST reconstruction, achieving an average RMSE below 1.3 K and SSIM of 0.939, outperforming six benchmark methods including STARFM, ESTARFM, and GAN-based approaches. Guo et al. (2025) proposed a physics-guided hierarchical transformer framework for SST forecasting and marine heatwave detection, integrating a U-shaped encoder–decoder with a temporal–spatial predictor module and a physics-constrained branch grounded in the mixed-layer heat budget equation, significantly outperforming baselines including ConvLSTM, DeepONet, and FNO.

3.2.4. Classical Supervised Methods

Despite the dominance of deep learning, classical supervised ML methods remained important, particularly in LST-related applications where interpretability is prioritised (Khalil and Kumar, 2025):

- Random Forest (RF): 15 full-text studies. Predominantly used for LST downscaling and UHI–LST relationship modelling.

- Support Vector Machine / Regression (SVM/SVR): 30 full-text studies. Frequently employed as a baseline comparator against deep learning methods. SVR consistently underperformed CNN and LSTM models in spatiotemporally complex tasks but remained competitive in data-scarce settings.
- Artificial Neural Networks (ANN / MLP): 17 full-text studies. Shallow multi-layer perceptrons appeared across all application domains as baseline models or as the final regression head within hybrid architectures (Shatnawi and Abu Qdais, 2019).

3.2.5. Physics-Informed and Hybrid Approaches

Physics-guided neural networks appeared in several full-text studies. Dong et al. (2026) proposed SSTFormer, a physics-guided deep learning framework for global SST forecasting across daily and monthly timescales, incorporating ocean current constraints to mitigate error accumulation in multi-step forecasting, achieving RMSE of 0.17 °C for daily and 0.60 °C for monthly forecasts. Shi et al. (2024) developed PANN, a physics-guided attention-based neural network for SST prediction in the East China Sea, which embedded PDE-based physical constraints via a dedicated physical constraint module and fused them with data-driven predictions through an attention fusion module. Transfer learning was employed in 7 full-text studies; Xu et al. (2024) demonstrated a TL approach that leveraged radiative transfer model simulations to improve LST retrieval from Landsat TOA reflective and emissive data, reducing dependence on land surface emissivity inputs and outperforming conventional ML, single-channel, and split-window algorithms.

Table 3. ML/DL architecture families identified across 93 full-text included studies. Percentages sum to >100% as many studies employed multiple architectures.

Architecture Family	n (full-text)	%	Primary Applications
Convolutional Neural Network (CNN/3D-CNN)	45	43.7%	LST retrieval, SST gap-filling, downscaling
LSTM / BiLSTM / ConvLSTM	34	33.0%	SST forecasting, LST time-series prediction
Support Vector Machine / SVR	30	29.1%	Baseline comparator, SST prediction, LST retrieval
Artificial Neural Network (ANN/MLP)	17	16.5%	All domains – baseline and regression head
Random Forest (RF)	15	14.6%	LST downscaling, UHI modelling
Generative Adversarial Network (GAN)	11	10.7%	SST / LST super-resolution, downscaling
Gradient Boosting (XGBoost/LightGBM)	9	8.7%	UHI, urban morphology–LST analysis, SST retrieval
Transformer / Vision Transformer (ViT)	13	14%	SST forecasting, eddy detection, LST retrieval
Encoder-Decoder / U-Net	7	6.8%	LST downscaling, SST reconstruction
Transfer Learning	7	6.8%	ECOSTRESS, Sentinel-3, cross-sensor adaptation

Architecture Family	n (full-text)	%	Primary Applications
Autoencoder / DINCAE	4	3.9%	SST / LST gap-filling
Physics-Informed NN (PINN)	3	2.9%	LST retrieval, SST prediction

3.3. Application Domain Coverage

Application domain data were available for **180 of 193 included studies**. The distribution reveals a strong concentration in oceanographic and land surface temperature applications, with a notable underrepresentation of emerging domains such as wildfire detection, permafrost monitoring, and evapotranspiration estimation.

3.3.1. Sea Surface Temperature Forecasting and Prediction (n ≈ 29, ~16%)

SST forecasting constituted the largest single application domain. Studies ranged from short-term (1–7 day) forecasts of coastal SST — where LSTM and ConvLSTM models demonstrated RMSE values typically in the range of 0.4–0.9°C in marginal and semi-enclosed seas — to medium-range monthly predictions, such as the EMD-LSTM model of Cai et al. (2025), which reported RMSE within 0.5°C for monthly average SST in the South China Sea. At the longer end of the spectrum, Krestenitis et al. (2024) introduced a deep learning framework for interim-future (1-year-ahead) daily SST forecasting over the northeastern Mediterranean Seas (Aegean, Ionian, and Cretan Seas), trained on 15 years of satellite data (2008–2022), demonstrating skill even for data more than two years beyond the training period. Subseasonal (3–12 month) predictions of global North Atlantic SST anomalies have also been approached using CNN and Transformer architectures.

A recurring finding across SST forecasting studies was the superior performance of spatiotemporal hybrid architectures over purely spatial or temporal models. Zhang et al. (2025) introduced a coupled Transformer–CNN network for SST forecasting, while Fu et al. (2024) combined LSTM with Transformer attention layers to improve SST prediction accuracy in regional seas. Jin et al. (2025) proposed the Decoupled Dynamic Spatial–Temporal Graph Neural Network (DDST-GNN) for SST prediction, exploiting non-Euclidean spatial dependencies in ocean thermal fields. Zhao et al. (2026) further extended graph-based approaches with Dynamic Graph Contrastive Learning for SST forecasting.

3.3.2. LST Retrieval from Satellite Thermal Sensors (n ≈ 46, ~25%)

LST retrieval studies applied ML/DL to derive land surface temperature directly from satellite-measured top-of-atmosphere radiances or brightness temperatures. CNNs dominated this domain. Ye et al. (2025) demonstrated end-to-end CNN-based LST retrieval that considers the thermal environment in sensor calibration, while Zhang et al. (2024b) integrated deep learning with split-window and TES algorithms for LST retrieval from Landsat 8. A notable methodological trend was multi-source sensor fusion: 11 of 23 full-text retrieval studies combined data from two or more thermal sensors (Han et al., 2024).

3.3.3. LST Downscaling and Spatiotemporal Fusion (n ≈ 23, ~13%)

LST downscaling studies addressed the persistent challenge of enhancing the spatial resolution of coarse thermal satellite products (typically MODIS at 1 km) to finer scales compatible with urban planning and ecosystem monitoring (30–90 m). Patil et al. (2024) explored scale-dependency effects in LST downscaling using a CNN-ELM hybrid, and Rogalski and Ilunga (2025) demonstrated Random Forest-based LST prediction from rainfall and elevation across South African dryland ecosystems. GAN-based super-resolution approaches (Jalbuena et al., 2025) consistently produced sharper thermal boundaries than CNN regression approaches, though at the cost of higher computational requirements.

3.3.4. LST and SST Gap-Filling and Cloud Reconstruction (n ≈ 20, ~11%)

Gap-filling studies have focused on reconstructing missing thermal data caused by cloud cover, sensor limitations, and orbital gaps. Deep learning approaches, such as the DINCAE autoencoder, have demonstrated improved performance over traditional EOF-based methods, reducing reconstruction error (e.g., RMSE from 0.46 to 0.38 °C) while better preserving spatiotemporal variability. El Azhary and Minaoui (2025) proposed EDDA-ConvLSTM for mesoscale SST reconstruction in the Mediterranean Sea with dual-attention mechanisms. Han et al. (2024) developed a time-continuous LST data fusion method for MODIS-Landsat spatiotemporal reconstruction that reduces cloud-induced data gaps in urban monitoring. Ge et al. (2025) predicted next-day 1-km canopy urban heat island maps using CNN-based spatiotemporal fusion, demonstrating the feasibility of near-real-time thermal urban monitoring.

3.3.5. Urban Heat Island (UHI) Modelling (n ≈ 5, ~3%)

UHI studies applied ML/DL to characterise spatial drivers of surface urban heat islands and forecast future UHI intensity. Singh et al. (2024) introduced the hybrid Xception-LSTM model for advanced urban heat island detection from satellite thermal imagery. Lin (2025) applied deep learning to integrate remote sensing data for urban heat island distribution observation in Chinese cities. Ge et al. (2025) developed a multi-block convolutional neural network (MBCNN) for predicting next-day 1-km canopy urban heat island intensity by integrating satellite- and ground-based observations. Gong et al. (2025) proposed a spatiotemporal deep learning approach to enhance air temperature estimation using only land surface temperature as input. Lu et al. (2025) investigated how 3D urban configurations shape UHI patterns in ECOSTRESS thermal data using Random Forest, identifying non-linear thresholds in building density and green space interactions.

3.3.6. Evapotranspiration, Agriculture, and Other Domains (n ≈ 4, ~2%)

Four full-text studies addressed thermal RS applications outside the core LST/SST domains. Jawad et al. (2025) developed a hybrid Penman–Monteith and deep learning (PMDL) framework for evapotranspiration estimation over drylands, achieving high accuracy ($R^2 > 0.85$; $KG_{Ess} > 0.78$) and outperforming widely used products such as MOD16A2, ECO3ETPTJPL, ECO3ETALEXI, and OpenET. Zhao et al. (2024) conducted a comprehensive assessment using four widely adopted machine learning models: deep forest (DF), deep neural networks (DNN), random forest (RF), and extreme gradient boosting (XGB). These models were employed to reconstruct ET products and fill gaps caused by cloud cover and model failures over the Heihe River Basin. The results demonstrated that all four methods performed effectively; however, the RF model showed the highest robustness. It achieved strong agreement with ground-based observations ($R = 0.73$) and successfully reconstructed missing values across the basin. The DNN and XGB models also performed well ($R > 0.70$), although residual gaps persisted in desert regions, particularly for XGB. The DF model was capable of filling gaps across the entire basin but exhibited lower consistency with in situ measurements ($R = 0.66$) and tended to underestimate ET values. Karimian et al. (2024) monitored spatiotemporal land use–land cover changes and their relationship to LST using deep learning in semiarid environments. Glacier and permafrost temperature monitoring studies applied SVM and ANN to Landsat TIRS and ASTER data over Southeast Asian highland regions.

Table 4. Distribution of included studies by application domain across all 193 studies. Studies without sufficient metadata for domain classification are listed as not categorised. Percentages are approximate given partial metadata availability.

Application Domain	n (all 193)	% (approx.)	Dominant Architecture	Key Sensors
LST Retrieval + Related	~46	~25%	CNN, Transformer, Hybrid	MODIS, ASTER, ECOSTRESS, Landsat TIRS
SST Forecasting / Prediction	~29	~16%	ConvLSTM, BiLSTM, Transformer, GNN	NOAA AVHRR, GOES, AMSR
LST Downscaling / Fusion	~23	~13%	CNN, GAN, Random Forest	MODIS, Landsat, Sentinel-3
LST / SST Gap-Filling	~20	~11%	CNN, Autoencoder, ConvLSTM	MODIS, Himawari AHI, AMSR
Urban Heat Island (UHI)	~5	~3%	Random Forest, XGBoost, LSTM, Xception	Landsat 8/9 TIRS, MODIS, ECOSTRESS
Evapotranspiration / Agriculture	~2	~1%	Deep NN, LSTM	ECOSTRESS, Landsat, MODIS
Permafrost / Glacier Temperature	~2	~1%	SVM, ANN	Landsat TIRS, ASTER
Not categorised (metadata incomplete)	~66	~34%	—	—

3.4. Sensor and Platform Usage

Thermal satellite sensor usage was identified from the 93 full-text studies. ASTER and Landsat TIRS emerged as the most frequently referenced platforms. The high frequency of ASTER (70%) likely reflects its use as a **validation reference** rather than primary input — ASTER's 90-m TIR resolution and laboratory-calibrated emissivity spectra have established it as the de facto ground-truth standard for validating coarser-resolution thermal products (Gillespie et al., 1998). ECOSTRESS was incorporated or compared in five recent studies (Jawad et al., 2025; Xu et al., 2024; Lu et al., 2025; Chang et al., 2024). Its relatively modest adoption likely reflects the complexity of accessing and processing ECOSTRESS data, as well as its irregular temporal sampling.

Table 5. Thermal satellite sensor usage across 93 full-text included studies. Studies frequently used multiple sensors; percentages sum to >100%.

Sensor / Platform	n (full-text)	%	Primary Application in Corpus
ASTER (Terra)	72	69.9%	LST retrieval, downscaling — spatial reference / validation
Landsat 8/9 TIRS	38	36.9%	LST retrieval, downscaling, UHI
NOAA AVHRR	36	35.0%	SST time series, historical gap-filling
MODIS (Terra/Aqua)	34	33.0%	LST products, downscaling input, SST
GOES-R ABI	17	16.5%	SST gap-filling, diurnal LST monitoring
Sentinel-3 SLSTR	17	16.5%	LST / SST retrieval, downscaling
AMSR / AMSR2	12	11.7%	All-sky SST, passive microwave LST
Meteosat / SEVIRI	9	8.7%	SST gap-filling, European LST
VIIRS (Suomi-NPP / NOAA-20)	9	8.7%	LST retrieval, SST continuation
ECOSTRESS (ISS)	5	4.9%	High-res LST retrieval, ET estimation
Himawari AHI	3	2.9%	Geostationary SST gap-filling

3.5. Performance Benchmarking

Quantitative performance metrics were extractable from the **93 full-text studies**: 48 studies reported RMSE and 16 studies reported R^2 . **Performance data from non-OA studies are not included in this synthesis**, as full methodological and quantitative details could not be verified from abstract metadata alone.

3.5.1. LST Retrieval

RMSE values reported in the literature vary widely depending on the modeling framework and evaluation methodology. Studies based on physical LST retrieval typically report errors in the range of ~0.2–3.3 K. In contrast, some machine learning studies report substantially lower RMSE values (<0.1), which reflect model evaluation on normalized or rescaled temperature variables and are therefore not directly comparable. Among physics-informed hybrid approaches, the MDK-DL framework — combining expert knowledge, radiation transfer modelling, and deep learning — achieved RMSE = 1.12 K validated against in situ data using three TIR remote sensing datasets (Gaofen, MODIS, and Fengyun) spanning high, medium, and low spatial resolutions. For hyperspectral data, an adaptive Fourier neural operator (AFNO)-transformer model was developed to retrieve LST from IASI observations by addressing channel redundancy through a linearized radiative transfer weighting scheme; validation against the AVHRR/MetOp LST product over Eastern Spain and North Africa demonstrated retrieval errors below 2.5 K, with broader temporal validation achieving errors below 3 K.

3.5.2. LST Downscaling

LST downscaling RMSE values among reviewed studies ranged from 0.886 K to 3.23 K. A hybrid ATC and 3D-CNN reconstruction method was validated on both Landsat 8 TIRS and MODIS datasets under cloud contamination and sensor failure conditions, with the proposed model achieving RMSE

of 0.96 K (Landsat) and 0.61 K (MODIS), outperforming baseline ATC methods that reached RMSE up to 3.23 K (Fu et al. 2022). Liang et al. (2023) proposed a GNNWR-based high-resolution LST downscaling method integrating NDVI, NDBI, DEM, and slope data, achieving a maximum R^2 of 0.974 and minimum RMSE of 0.896 K across four test areas with differing topography and seasons. Khedher et al. (2023) introduced a multimodal conditional GAN approach for high-resolution LST estimation integrating RGB orthophotography, elevation, and land cover, reporting RMSE of 1.5 K. Wang et al. (2024b) extended the GNNWR framework with Area-to-Point Kriging (GNNWRK) for nighttime LST downscaling, achieving minimum RMSE of 0.886 K and maximum Pearson correlation of 0.930 across four areas of varying landform and climate type, outperforming TsHARP, Random Forest, and Geographically Weighted Regression benchmarks.

3.5.3. SST Forecasting

SST forecasting RMSE values across reviewed studies ranged from 0.084 K to 1.40 K, reflecting strong variation across model types, forecast horizons, and ocean regions. Xu et al. (2023) proposed multi-region encoder-decoder LSTM models for global short-term SST prediction, achieving RMSE of 0.2712–0.6487 K over ten consecutive forecast days. Jiao et al. (2024) applied LSTM to coastal SST forecasting across 14 stations along the Bohai Sea and Yellow Sea, achieving 1-day RMSE of 0.28 K – an average 78% reduction in error compared to numerical models. Xie et al. (2024) developed a 3D U-Net model integrating multi-source data for South China Sea SST forecasting, achieving RMSE < 0.5 K across lead times of 1–30 days. Ji et al. (2024) found that incorporating atmospheric forcing variables into a SA-ConvLSTM model improved East China Sea SST prediction skill by up to 10.75% in RMSE, though errors remained highest in summer.

3.5.4. SST Gap-Filling and Reconstruction

SST gap-filling studies reported RMSE values ranging from 0.46 K to 2.07 K across reviewed papers. A CNN-based approach integrating geostationary Advanced Himawari Imager (AHI) with Advanced Technology Microwave Sounder (ATMS) data achieved RMSE = 2.07 K and MAE = 1.22 K for all-weather SST retrieval, substantially outperforming unimodal AHI-only (RMSE = 3.45 K) and ATMS-only (RMSE = 2.64 K) baselines (Fan et al., 2025). Putra and Hsu (2024) proposed a Double U-Net model pretrained on Level-4 SST images for gap-filling cloud-obscured SST near Taiwan, achieving RMSE = 1.12 K and outperforming Data Interpolating Empirical Orthogonal Functions (DINEOFs) in spatial detail and buoy-observation accuracy. A convolutional auto-encoder approach (DINCAE) applied to a 25-year AVHRR SST time series reconstructed cloud-masked data with RMSE = 0.46 K, handling variable-accuracy missing data through likelihood-maximisation training (Barth et al., 2020).

Table 6. Summary of reported performance metrics by application domain across 93 full-text included studies, with representative study citations. Non-OA studies are excluded as performance data could not be verified from abstract metadata.

Domain	Metric	Best Reported	Typical Range	Representative Studies
LST Retrieval	RMSE (K)	0.2	0.2–3.29	Khedher et al. (2024); He et al. (2023)
LST Retrieval	R^2	0.99	0.69–0.99	Ji et al. (2022); Wang et al. (2024)
LST Downscaling	RMSE (K)	0.886	0.886–3.23	Liang et al. (2023); Wang et al. (2024)

Domain	Metric	Best Reported	Typical Range	Representative Studies
LST Downscaling	R ²	0.97	0.58–0.97	Liang et al. (2023); Fu et al. (2022)
SST Forecasting	RMSE (K)	0.084	0.084–1.40	Liang et al. (2023); Xu et al. (2023)
SST Forecasting	R ²	0.996	0.956–0.996	Li et al. (2025); Shi et al. (2026)
SST Gap-Filling	RMSE (K)	0.463	0.46–2.07	Barth et al. (2020); Putra & Hsu (2024)
SST Gap-Filling	MAE (K)	0.023	0.023–1.22	Goh et al. (2024); Fan et al. (2025)
UHI / LST Mapping	RMSE (K)	1.042	1.042–1.65	Jeong et al. (2020); Kporha (2026)
UHI / LST Mapping	R ²	0.966	0.966	Jeong et al. (2020)

3.6. Methodological Quality and Open Science Practices

Methodological reporting quality varied substantially across the 93 full-text studies. Common limitations included: (i) single-site or single-region validation without spatial cross-validation, present in ~60% of LST downscaling and UHI studies; (ii) absence of baseline comparisons against physics-based algorithms in 31% of LST retrieval studies; (iii) non-standardised performance metrics, complicating cross-study comparison; and (iv) data leakage risk in spatiotemporal settings where temporal overlap between training and test sets was not always explicitly excluded — for example, one LST retrieval study applied a random pixel-level division of the full time series without chronological separation (Tan et al., 2019), and 22 of 75 LST/UHI studies provided no information on split strategy, making temporal exclusion unverifiable. Open science practices were inconsistently adopted. Code availability was explicitly stated in fewer than 5% of included full-text studies, with the majority of open-source repositories associated with studies published in *Remote Sensing*, *ISPRS Journal of Photogrammetry and Remote Sensing*, or *IEEE Transactions on Geoscience and Remote Sensing*. Training data availability was similarly limited: while most studies relied on publicly accessible satellite archives (MODIS, Landsat, ECOSTRESS, NOAA AVHRR), the specific preprocessed training datasets were rarely deposited in open repositories, reducing reproducibility.

4. Discussion

The synthesis of 193 studies presented in Section 3 reveals several overarching findings that merit critical discussion. The following subsections address each in turn, contextualising findings within the broader ML/DL remote sensing literature (Zhu et al., 2017; LeCun et al., 2015).

4.1. Architectural Conservatism and the Transformer Transition

The dominance of CNN and LSTM architectures — collectively present in approximately 63% of full-text included studies — reflects the maturity of these frameworks for spatiotemporal thermal RS modelling, but also indicates a potential conservatism in the community's adoption of newer

architectures. Transformer-based models, despite their demonstrated utility in long-range dependency modelling relevant to climate-scale SST prediction and global LST retrieval from hyperspectral sounders (e.g., the AFNO-transformer applied to IASI retrievals; Lan and Yang, 2023), remained a minority of the corpus (13 full-text studies, 14%). This gap is likely to close rapidly: the inclusion of five transformer-based studies published in 2025–2026 alone (Fan et al., 2025; Liang et al., 2025; Ji et al., 2025; Wang et al., 2026; Shi et al., 2026) suggests an accelerating uptake trajectory consistent with the broader remote sensing community's transition to transformer architectures.

The emergence of frontier architectures was detectable in the 2025–2026 literature. The HiT_DS hierarchical Transformer of Wang et al. (2026) integrates physics-informed loss functions with gradient-aware attention for oceanographic downscaling, exemplifying the convergence of physical constraints with attention-based modelling. Foundation models — large pre-trained geospatial models such as SatMAE and Prithvi — were not represented in the confirmed corpus, and formal benchmark evaluations for transformer architectures on thermal RS tasks remain sparse.

4.2. Application Domain Imbalance

The application domain imbalance is striking. SST forecasting, LST retrieval, LST downscaling, and LST gap-filling together account for the vast majority of included studies, while societally critical applications — wildfire thermal anomaly detection, evapotranspiration estimation, permafrost temperature monitoring, glacier calving thermal dynamics — are severely underrepresented. This concentration likely reflects data availability and methodological tractability: SST forecasting and LST downscaling benefit from long, regular satellite records (AVHRR extends to the 1980s; MODIS to 2000) (Reynolds et al., 2002), while wildfire detection and permafrost monitoring require event-triggered or site-specific data collection.

The UHI domain, well-represented in traditional remote sensing literature (Zhou et al., 2019; Rasul et al., 2017), accounted for only approximately 3% of the 193 included studies — a finding that likely reflects our strict inclusion criteria requiring ML/DL as a core analytical component. Studies using standard NDVI–LST regression without an ML layer were correctly excluded, but this may have inadvertently reduced the UHI count.

4.3. The Physics–Data Integration Gap

Only 3 of 93 full-text studies employed physics-informed neural networks, despite compelling theoretical arguments for their advantages in thermal RS: the radiative transfer equation governing satellite LST retrieval is well-characterised (Wan and Dozier, 1996; Gillespie et al., 1998), physically based formulations such as the Penman–Monteith equation provide a robust foundation for hybrid evapotranspiration models by constraining learning with surface energy balance principles (Jawad et al., 2025), and ocean heat transport dynamics constrain SST variability. The MDK-DL framework demonstrated that physics embedding can improve cross-sensor generalisation in LST retrieval across multiple TIR sensors (Wang et al., 2021). Dong et al. (2026) and Shi et al. (2024) provide compelling recent examples of physics-guided DL for SST forecasting and prediction respectively, suggesting that the barrier to PINN adoption is beginning to lower.

4.4. Open Science Deficit

The open science deficit identified across the corpus represents a structural reproducibility challenge. Without deposited code and preprocessed training data, the results reported across these studies cannot be independently verified. This issue is particularly acute for operational applications — LST gap-filling (Han et al., 2024; El Azhary and Minaoui (2025), SST forecasting (Krestenitis et al., 2024; Song et al., 2024) — where reproducibility of model training pipelines is a prerequisite for operational deployment. The observation that code availability remains low across the corpus suggests that voluntary adoption of open science practices is insufficient, and that journal-level

enforcement of data and code sharing policies may be necessary to improve reproducibility in the field.

4.5. Implications of the Data Access Stratification

The stratified inclusion of 93 full-text and 100 metadata-only studies is a methodological transparency choice consistent with PRISMA 2020 best practices (Page et al., 2021). The 100 non-OA studies are not absent from this review — they are present in the publication trend analysis, architecture classification (where identifiable from metadata), and application domain mapping. Their exclusion from performance benchmarking is a data quality decision, not a coverage decision. Readers should note that performance metrics in Section 3.5 reflect only the 93 full-text studies and may therefore underrepresent the full distribution of model performance across the field.

5. Research Gaps and Future Directions

Based on the systematic synthesis of 193 included studies, six priority research gaps are identified:

5.1. Transformer and Foundation Model Adoption for Thermal RS

While Hu et al. (2025) and Guo et al. (2025) have demonstrated the potential of transformer-based architectures for thermal spatiotemporal fusion and SST forecasting respectively, benchmarking these approaches against CNN and LSTM baselines across the full range of thermal RS tasks identified in this review remains a high-priority research direction.

5.2. Physics-Informed Neural Networks for Operational LST Retrieval

The demonstrated gains of MDK-DL, the AFNO-transformer, Dong et al. (2026), and Shi et al. (2024) should be extended to a wider range of sensors (ECOSTRESS, Sentinel-3 SLSTR, GOES ABI) and atmospheric conditions, with rigorous comparison against operational split-window (Wan and Dozier, 1996) and TES (Gillespie et al., 1998) algorithms. The development of shared, differentiable implementations of the radiative transfer equation for common TIR sensor systems would substantially lower the entry barrier for PINN adoption.

5.3. Underrepresented Application Domains

Wildfire thermal anomaly detection, evapotranspiration mapping at field scale (Jawad et al., 2025), permafrost active layer temperature monitoring, and glacier calving front thermal dynamics are substantially underserved by ML/DL approaches relative to their scientific and societal importance. Targeted benchmark dataset development — analogous to the widely used MODIS fire products that already provide labelled thermal anomaly training data — would catalyse ML/DL adoption in these critical areas.

5.4. Standardised Benchmarking Datasets

The absence of community-standard benchmark datasets for thermal RS — analogous to ImageNet for computer vision — prevents systematic comparison of architectures across studies and impedes identification of genuine performance advances. Community initiatives to establish standardised train/test splits for MODIS LST downscaling, ECOSTRESS gap-filling, and SST forecasting benchmarks would be a high-value contribution, comparable in impact to the community benchmarking initiatives that have accelerated progress in oceanographic and atmospheric ML more broadly.

5.5. Spatiotemporal Generalisation and Cross-Site Validation

The overwhelming majority of included studies validated ML/DL models within the same geographic region used for training (Tanoori et al., 2024; Karimian et al., 2024; Rogalski and Ilunga, 2025). Systematic evaluation of model transferability across climate zones, biomes, and sensor systems is needed before operational deployment can be justified. The adoption of leave-one-region-out cross-validation frameworks — standard in ecological modelling but rare in thermal RS ML studies — would provide more rigorous estimates of generalisation performance.

5.6. Open Science Infrastructure

The systematic adoption of code and data sharing practices — facilitated by platforms such as GitHub, Zenodo, and Google Earth Engine — is necessary to enable reproducibility and community-driven improvement of ML/DL thermal RS models. Journal policies encouraging code and data availability statements, such as those adopted by Remote Sensing of Environment and ISPRS Journal of Photogrammetry and Remote Sensing, should be strengthened and extended across all target publication venues. Google Scholar retrieval may have partial reproducibility limitations due to dynamic indexing and access constraints.

6. Conclusions

This PRISMA 2020-compliant systematic review synthesised 193 peer-reviewed studies applying machine learning and deep learning to thermal remote sensing published between January 2018 and March 2026 — the most comprehensive synthesis of ML/DL in this domain to date. The following principal conclusions are drawn:

- Publication output grew exponentially over the review period, from 6 studies in 2018–2019 to 62 studies in 2024 and 65 in 2025 alone — with a further 19 already published by March 2026 — driven by converging advances in satellite thermal sensor constellations (Fisher et al., 2020; Xu et al., 2024), open-source deep learning frameworks, and open-access data archives including Landsat Collection 2, MODIS Version 6.1, and ECOSTRESS.
- CNN and LSTM architectures dominated the corpus, collectively present in approximately 63% of full-text included studies (CNN: 43.7%; LSTM/BiLSTM/ConvLSTM: 33.0%), with SVM/SVR constituting the third most common method (29.1%). Transformer-based models represent a rapidly growing minority (14%; 13 full-text studies), with five transformer studies published in 2025–2026 alone, suggesting architectural parity with CNNs is approaching. Foundation model adoption (e.g., SatMAE, Prithvi) was not yet represented in the confirmed corpus.
- Application domains were heavily concentrated in LST retrieval (~25%), SST forecasting (~16%), LST downscaling (~13%), and LST/SST gap-filling (~11%). Wildfire detection, evapotranspiration estimation (Jawad et al., 2025), permafrost monitoring, and glacier thermal dynamics — collectively accounting for fewer than 2% of included studies — are critically underserved relative to their scientific and societal importance.
- Physics-informed neural networks (PINNs) appeared in only 3 of 93 full-text studies (2.9%) despite demonstrating substantive accuracy gains over purely data-driven approaches: Dong et al. (2026) achieved RMSE of 0.17 °C for daily global SST forecasts; Shi et al. (2024) embedded PDE-based physical constraints into an attention-based SST prediction network; and the MDK-DL framework improved cross-sensor LST retrieval generalisation across multiple TIR sensors. Their broader adoption — facilitated by shared, differentiable implementations of the radiative transfer equation — represents the single most impactful near-term methodological opportunity identified by this review.

- Open science practices were critically underadopted: code availability was explicitly stated in fewer than 5% of full-text included studies, and preprocessed training datasets were rarely deposited in open repositories despite most studies relying on publicly accessible satellite archives. Methodological quality deficits were widespread — approximately 60% of LST downscaling and UHI studies relied on single-site validation without spatial cross-validation, 31% of LST retrieval studies omitted physics-based baseline comparisons, and temporal data leakage risk was unverifiable in 22 of 75 LST/UHI studies.
- The stratified data access approach adopted — full quantitative synthesis for 93 open-access studies, descriptive analysis for 100 metadata-only studies — is methodologically transparent and consistent with PRISMA 2020 standards (Page et al., 2021). Future systematic updates conducted with broader institutional access would strengthen the completeness of the performance benchmarking synthesis.

Collectively, these findings establish that ML/DL thermal RS has matured from a niche research frontier into a mainstream Earth observation methodology. The field now faces three interlocking challenges: methodological consolidation — standardising community benchmark datasets for LST downscaling, SST forecasting, and gap-filling; improving open science compliance and reproducibility; and deepening physics–data integration through PINNs and hybrid frameworks. Simultaneously, systematic thematic expansion towards critically underserved domains — wildfire thermal anomaly detection, field-scale evapotranspiration, permafrost active layer monitoring, and glacier calving thermal dynamics — and rigorous cross-site spatiotemporal generalisation testing are prerequisites for the field to realise its full scientific and societal potential.

Appendix A: 2020. Checklist Summary

PRISMA 2020 Item	Section in This Review	Status
Title: identify as systematic review	Title Page	✓ Compliant
Abstract: structured with background, methods, results, conclusions	Abstract	✓ Compliant
Introduction: rationale and objectives	Section 1	✓ Compliant
Eligibility criteria: inclusion and exclusion criteria	Section 2.2	✓ Compliant
Information sources: databases and search dates	Section 2.1	✓ Compliant
Search strategy: full keyword blocks provided	Sections 2.1.1–2.1.2	✓ Compliant
Selection process: documented pipeline	Section 2.3	✓ Compliant
Data collection: extraction template	Section 2.5	✓ Compliant

Data items: list of variables	Section 2.5	✓ Compliant
Study risk of bias assessment	Section 2.7	✓ Compliant
Effect measures: narrative synthesis stated	Section 2.8	✓ Compliant
Synthesis methods: heterogeneity addressed	Section 2.8	✓ Compliant
Study selection results: PRISMA flowchart	Table 1	✓ Compliant
Study characteristics: per-study supplementary file	Appendix B	✓ Compliant
Risk of bias: reported by domain	Section 3.6	✓ Compliant
Certainty of evidence: addressed in discussion	Section 4.5	✓ Compliant
Data access limitation: transparently reported	Section 2.4	✓ Compliant

Appendix B: Scope of Included Studies

The full reference list of all 193 included studies is provided as a supplementary Excel file (final_included_list.xlsx) accompanying this manuscript. The spreadsheet includes, for each study: DOI, title, authors, year, journal, abstract (where available), ML method, application domain, data source, PDF availability status, and screening rationale. Studies are organised into two sheets: (i) From PDFs (93 open-access full-text studies) and (ii) From Non-OA Screening (100 metadata-only studies). This structured supplementary file constitutes the complete data extraction record and enables replication of all descriptive statistics reported in Section 3.

References

- Barth, A., Alvera-Azcárate, A., Licer, M., & Beckers, J. M. (2020). DINCAE 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geoscientific Model Development*, 13(3), 1609–1622. <https://doi.org/10.5194/gmd-13-1609-2020>
- Cai, Z., Ren, K., Li, B., Guo, W., Hui, G., & Zhang, X. (2025). Study on sea surface temperature prediction of South China Sea based on EMD-LSTM. In *2025 IEEE 2nd International Conference on Big Data Science and Engineering (ICBDSE)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ICBDSE65491.2025.11220045>
- Chang, Y., Cao, Y., & Weng, Q. (2024, July). Generating hourly 70-m land surface temperature from GOES-R observations: A comparison of statistical downscaling and deep learning methods. In *IGARSS 2024–2024 IEEE International Geoscience and Remote Sensing Symposium* (pp. 918–920). IEEE. <https://doi.org/10.1109/IGARSS53475.2024.10640401>

- Chen, K., Tian, M., Zhang, J., Xu, X., & Yuan, L. (2023). Evaluating the seasonal effects of building form and street view indicators on street-level land surface temperature using random forest regression. *Building and Environment*, 228, 110884. <https://doi.org/10.1016/j.buildenv.2023.110884>
- Choe, Y. J., & Yom, J. H. (2020). Improving accuracy of land surface temperature prediction model based on deep-learning. *Spatial Information Research*, 28(3), 377–382. <https://doi.org/10.1007/s41324-019-00299-5>
- Dong, S., Li, Y., Yin, X., Xu, Q., & Mao, P. (2026). Physics-guided deep learning for global sea surface temperature forecasting: Balancing accuracy and stability across timescales. *Geoscience Frontiers*, 102255. <https://doi.org/10.1016/j.gsf.2026.102255>
- El Azhary, F. Z., & Minaoui, K. (2025). EDDA-ConvLSTM: Encoder-decoder dual attention ConvLSTM for Moroccan coastal sea surface temperature prediction. *IEEE Geoscience and Remote Sensing Letters*. <https://doi.org/10.1109/LGRS.2025.3551623>
- Fan, D., Yang, X., He, H., He, H., & Fu, B. (2025). Bridging the cloud gap: AHI/ATMS synergy through CNN feature fusion for all-weather SST retrieval. *International Journal of Applied Earth Observation and Geoinformation*, 144, 104887. <https://doi.org/10.1016/j.jag.2025.104887>
- Fisher, J. B., Lee, B., Purdy, A. J., Halverson, G. H., Dohlen, M. B., Cawse-Nicholson, K., & Hook, S. J. (2020). ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the International Space Station. *Water Resources Research*, 56(4), e2019WR026058. <https://doi.org/10.1029/2019WR026058>
- Fu, H., Shao, Z., Fu, P., Huang, X., Cheng, T., & Fan, Y. (2022). Combining ATC and 3D-CNN for reconstructing spatially and temporally continuous land surface temperature. *International Journal of Applied Earth Observation and Geoinformation*, 108, 102733. <https://doi.org/10.1016/j.jag.2022.102733>
- Fu, Y., Song, J., Guo, J., Fu, Y., & Cai, Y. (2024). Prediction and analysis of sea surface temperature based on LSTM-transformer model. *Regional Studies in Marine Science*, 78, 103726. <https://doi.org/10.1016/j.rsma.2024.103726>
- Ge, S., Zhan, W., Li, J., Li, L., Dong, P., Li, X., & Gao, Y. (2025). Prediction of next-day 1-km canopy urban heat island by integrating a multi-block convolutional neural network with satellite- and ground-based observations. *Sustainable Cities and Society*, 106701. <https://doi.org/10.1016/j.scs.2025.106701>
- Gillespie, A., Rokugawa, S., Matsunaga, T., Cothorn, J. S., Hook, S., & Kahle, A. B. (1998). A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), 1113–1126. <https://doi.org/10.1109/36.508406>
- Goh, E., Yepremyan, A., Wang, J., & Wilson, B. (2024). MAESSTRO: Masked autoencoders for sea surface temperature reconstruction under occlusion. *Ocean Science*, 20(5), 1309–1323. <https://doi.org/10.5194/os-20-1309-2024>
- Gong, P., Lan, L., Guo, C., & Guo, R. (2025). A spatiotemporal deep learning approach to enhance air temperature estimation based on the sole input of land surface temperature. *Journal of Cleaner Production*, 532, 146887. <https://doi.org/10.1016/j.jclepro.2025.146887>
- Guo, Y., Song, J., Cao, X., & Leng, H. (2025). A physics-guided hierarchical transformer framework for sea surface temperature forecasting and marine heatwave detection. *IEEE Geoscience and Remote Sensing Letters*, 23, 1–5. <https://doi.org/10.1109/LGRS.2025.3643727>
- Han, J., Fang, S., Mi, Q., Wang, X., Yu, Y., Zhuo, W., & Peng, X. (2024). A time-continuous land surface temperature (LST) data fusion approach based on deep learning with microwave remote sensing and high-density ground truth observations. *Science of the Total Environment*, 914, 169992. <https://doi.org/10.1016/j.scitotenv.2024.169992>

- He, Q., Cao, J., Saide, P. E., Ye, T., & Wang, W. (2024). Unraveling the influence of satellite-observed land surface temperature on high-resolution mapping of ground-level ozone using interpretable machine learning. *Environmental Science & Technology*, 58(36), 15938–15948. <https://doi.org/10.1021/acs.est.4c02926>
- He, T., Liu, F., Wang, A., Zhang, Z., & others. (2023). Estimating monthly surface air temperature using MODIS LST data and an artificial neural network in the Loess Plateau, China. *Chinese Geographical Science*, 33, 751–763. <https://doi.org/10.1007/s11769-023-1370-0>
- He, X., Liu, X., & Ru, C. (2025). Evaluating spatial representativeness across multiple scales for LST downscaling using random forest. *IEEE Transactions on Geoscience and Remote Sensing*. <https://doi.org/10.1109/TGRS.2025.3570685>
- Hu, P., Pan, X., Yang, Y., Dai, Y., & Chen, Y. (2025). A two-stage hierarchical spatiotemporal fusion network for land surface temperature with transformer. *IEEE Transactions on Geoscience and Remote Sensing*. <https://doi.org/10.1109/TGRS.2025.3552577>
- Jalbuena, R., Yee, J.-J., Yun, G. Y., & Raj, S. (2025). Addressing landcover bias in spatial downscaling of MODIS land surface temperature using generative adversarial network-based regression model (RGAN). *Advances in Space Research*, 76(6), 3445–3464. <https://doi.org/10.1016/j.asr.2025.07.024>
- Jawad, M., Behrangi, A., Farmani, M. A., Qiu, Y., Sohi, H. Y., Gupta, A., & Niu, G. Y. (2025). Improved evapotranspiration estimation using the Penman-Monteith equation with a deep learning (DNN) model over the dry southwestern US: Comparison with ECOSTRESS, MODIS, and OpenET. *Journal of Hydrology*, 660, 133460. <https://doi.org/10.1016/j.jhydrol.2025.133460>
- Ji, C., Xu, W., Zheng, X., Ahmed, Y., Jamal, S. A., Imam, F., & Kucher, D. E. (2025). Transformer-based deep learning for mesoscale eddy detection in sea surface temperature maps. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/JSTARS.2025.3574004>
- Ji, Q., Jia, X., Jiang, L., Xie, M., Meng, Z., Wang, Y., & Lin, X. (2024). Contribution of atmospheric factors in predicting sea surface temperature in the East China Sea using the random forest and SA-ConvLSTM model. *Atmosphere*, 15(6), 670. <https://doi.org/10.3390/atmos15060670>
- Jia, H., Yang, D., Deng, W., Wei, Q., & Jiang, W. (2021). Predicting land surface temperature with geographically weighted regression and deep learning. *WIREs Data Mining and Knowledge Discovery*, 11, e1396. <https://doi.org/10.1002/widm.1396>
- Jia, W., Guan, S., & Xue, Y. (2024). TL-iTransformer: Revolutionizing sea surface temperature prediction through iTransformer and transfer learning. *Earth Science Informatics*, 17, 4847–4857. <https://doi.org/10.1007/s12145-024-01436-x>
- Jiao, Y., Li, G., Zhao, P., Chen, X., Cao, Y., Liu, G., & Ji, C. (2024). Construction of sea surface temperature forecasting model for Bohai Sea and Yellow Sea coastal stations based on long short-time memory neural network. *Water*, 16(16), 2307. <https://doi.org/10.3390/w16162307>
- Jin, K., He, X., Yang, J., Jin, R., & Fu, F. (2025). Decoupled dynamic spatial-temporal graph neural network for sea surface temperature prediction. *IEEE Geoscience and Remote Sensing Letters*. <https://doi.org/10.1109/LGRS.2025.3558739>
- Karimian, R., Rangzan, K., Karimi, D., & others. (2024). Spatiotemporal monitoring of land use–land cover and its relationship with land surface temperature changes based on remote sensing, GIS, and deep learning. *Journal of the Indian Society of Remote Sensing*, 52, 2461–2481. <https://doi.org/10.1007/s12524-024-01958-3>
- Khalil, M., & Kumar, J. S. (2025). Predictive modeling of land surface temperature dynamics in Damascus, Syria using Google Earth Engine: A remote sensing and random forest approach. *Earth Science Informatics*, 18, 328. <https://doi.org/10.1007/s12145-025-01844-7>

- Khalilabadi, M. R., & Sadati, S. (2025). A new hybrid deep learning-based method for sea surface temperature prediction. *Ships and Offshore Structures*, 1–8. <https://doi.org/10.1080/17445302.2025.2497554>
- Khanifar, J., & Khademalrasoul, A. (2022). Modeling of land surface temperature–multiscale curvatures relationship using XGBoost algorithm (Case study: Southwestern Iran). *International Journal of Environmental Science and Technology*, 19, 11763–11774. <https://doi.org/10.1007/s13762-022-04409-z>
- Khedher, I., Favreau, J. M., Miguet, S., & Gesquière, G. (2024). A multimodal deep learning approach for high-resolution land surface temperature estimation. In S. M. Thampi, J. Hu, A. K. Das, J. Mathew, & S. Tripathi (Eds.), *Applied soft computing and communication networks: ACN 2023. Lecture Notes in Networks and Systems* (Vol. 966). Springer. https://doi.org/10.1007/978-981-97-2004-0_26
- Kporha, V. K., Fox, D. M., Banitalebi, M., Bouroubi, Y., & Fournier, R. (2026). Comparing daily and 8-day MODIS land surface temperature data for urban heat island assessment using random forest modeling in data-limited regions. *Remote Sensing Applications: Society and Environment*, 101904. <https://doi.org/10.1016/j.rsase.2026.101904>
- Krestenitis, M., Androulidakis, Y., & Krestenitis, Y. (2024). Deep learning-based forecasting of sea surface temperature in the interim future: Application over the Aegean, Ionian, and Cretan Seas (NE Mediterranean Sea). *Ocean Dynamics*, 74, 149–168. <https://doi.org/10.1007/s10236-023-01595-3>
- Lan, X., & Yang, X. (2023). Retrieving land surface temperatures from IASI hyperspectral thermal infrared data using an AFNO-transformer model. *Optics Express*, 31(24), 40249–40260. <https://doi.org/10.1364/OE.502559>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, J. (2024). Evaluating the correlation between impacting factors and land surface temperature via spatial regression model and random forest. *The Professional Geographer*, 76(6), 740–755. <https://doi.org/10.1080/00330124.2024.2398243>
- Li, J., Yan, Y., & Stouffs, R. (2025). Simulating land surface temperature impacts of proposed land use and land cover plans using an integrated deep neural network approach. *Energy and Buildings*, 332, 115437. <https://doi.org/10.1016/j.enbuild.2025.115437>
- Li, R., Zhang, Y., & Chen, X. (2026). Urban land surface temperature prediction by integrating LSTM and geospatial information: A case study of Kunming, China. *Journal of Cleaner Production*, 548, 147856.
- Li, Z. L., Tang, B. H., Wu, H., Ren, H., Yan, G., Wan, Z., & Sobrino, J. A. (2013). Satellite-derived land surface temperature: Current status and perspectives. *Remote Sensing of Environment*, 131, 14–37. <https://doi.org/10.1016/j.rse.2012.12.008>
- Liang, M., Zhang, L., Wu, S., Zhu, Y., Dai, Z., Wang, Y., & Du, Z. (2023). A high-resolution land surface temperature downscaling method based on geographically weighted neural network regression. *Remote Sensing*, 15(7), 1740. <https://doi.org/10.3390/rs15071740>
- Liang, Q., Pan, Z., Zhang, X., Ye, Z., Zhao, Z., & Guo, S. (2025). Vision transformer applied for time series based prediction of sea surface temperature. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3595579>
- Lin, H. (2025). Urban heat island distribution observation by integrating remote sensing technology and deep learning. *International Journal of Image and Data Fusion*, 16(1). <https://doi.org/10.1080/19479832.2024.2354754>
- Lu, Y., Cheng, X., Yang, Y., & Wang, H. (2025). Thresholds and synergies: How 3D urban configurations shape thermal environment optimization in Chengdu's core—a multidimensional analysis with ECOSTRESS and random forest. *Energy and Buildings*, 116815. <https://doi.org/10.1016/j.enbuild.2025.116815>
- Norman, J. M., & Becker, F. (1995). Terminology in thermal infrared remote sensing of natural surfaces. *Agricultural and Forest Meteorology*, 77(3–4), 153–166. [https://doi.org/10.1016/0168-1923\(95\)02259-Z](https://doi.org/10.1016/0168-1923(95)02259-Z)

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Patil, J., Maithani, S., & Sharma, S. K. (2024). Exploring effect of scale dependency in LST downscaling using convolution neural network–extreme learning machine (CNN–ELM). *Earth Science Informatics*. <https://doi.org/10.1007/s12145-024-01239-0>
- Putra, D. P., & Hsu, P. C. (2024). Leveraging transfer learning and U-Net method for improved gap filling in Himawari sea surface temperature data adjacent to Taiwan. *ISPRS International Journal of Geo-Information*, 13(5), 162. <https://doi.org/10.3390/ijgi13050162>
- Qi, J. F., Xie, B. W., Yang, D. Z., Li, D. L., & Yin, B. S. (2024). Deep learning-driven forecasting of daily sea surface temperatures in the South China Sea. In *2024 Photonics & Electromagnetics Research Symposium (PIERS)* (pp. 1–6). IEEE. <https://doi.org/10.1109/PIERS62282.2024.10617964>
- Qi, T., Tang, S., & Ye, H. (2025). RVDCU-Net: A modular deep learning framework for spatiotemporal reconstruction of satellite-derived sea surface temperature. *Marine Geodesy*, advance online publication. <https://doi.org/10.1080/01490419.2025.2583086>
- Rani, B. S., Narayanan, D. S., Jasmine, M. B. S., Krishnamoorthy, N. R., Raman, M. R., Lotliker, A. A., & Chatterjee, A. (2025). Deep learning for sea surface temperature prediction in the Indian Ocean: A comparative study using 1D-CNN, LSTM, 1DCNN-LSTM and attention-based 1DCNN-LSTM architectures. *Dynamics of Atmospheres and Oceans*. <https://doi.org/10.1016/j.dynatmoce.2025.101632>
- Rasul, A., Balzter, H., Smith, C., Remedios, J., Adamu, B., Sobrino, J. A., & Weng, Q. (2017). A review on remote sensing of urban heat and cool islands. *Land*, 6(2), 38. <https://doi.org/10.3390/land6020038>
- Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, 15(13), 1609–1625. [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AIISAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2)
- Rogalski, S., & Ilunga, M. (2025). Predicting land surface temperature from rainfall and elevation remotely sensed data using random forest machine learning in Google Earth Engine. In *Proceedings of the International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2025)* (pp. 119–126). <https://www.iiis.org/CDs2025/CD2025Spring/papers/ZA794JM.pdf>
- Shatnawi, N., & Abu Qdais, H. (2019). Mapping urban land surface temperature using remote sensing and ANN. *International Journal of Remote Sensing*, 40(10), 3968–3983. <https://doi.org/10.1080/01431161.2018.1557792>
- Shi, B., Feng, L., He, H., Hao, Y., & Peng, Y. (2024). A physics-guided attention-based neural network for sea surface temperature prediction. *IEEE Transactions on Geoscience and Remote Sensing*. <https://doi.org/10.1109/TGRS.2024.3457039>
- Shi, B., Lin, H., Zhu, Y., He, H., Tan, Q., Peng, Y., & Liu, J. (2026). CA-ConvLSTM: A cross-attention-based deep learning framework for joint prediction of sea surface temperature and sea level anomaly. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/JSTARS.2026.3662796>
- Singh, A. K., Deepika, C. L. N., Shahnaz, K. V., & others. (2025). Hybrid Xception-LSTM model for remote sensing: Advanced urban heat island and land use analysis. *Remote Sensing in Earth Systems Sciences*, 8, 132–144. <https://doi.org/10.1007/s41976-024-00182-4>
- Song, D., Dai, S., Li, W., & Tong, T. (2024). STVformer: A spatial-temporal-variable transformer with auxiliary data for sea surface temperature forecasting. *Applied Ocean Research*. <https://doi.org/10.1016/j.apor.2024.104218>

- Sun, Z., & Wang, Y. (2025). A coordination attention residual U-Net model for enhanced short and mid-term sea surface temperature prediction. *Environmental Modelling & Software*, 183, 106251. <https://doi.org/10.1016/j.envsoft.2024.106251>
- Sundar, G., Patchaiammal, P., Lakkshmanan, A., & others. (2024). Comprehensive earth system analysis for advancing community resilience: Integrating remote sensing with hybrid GAN-RNN techniques. *Remote Sensing in Earth Systems Sciences*, 7, 788–803. <https://doi.org/10.1007/s41976-024-00164-6>
- Tan, J., NourEldeen, N., Mao, K., Shi, J., Li, Z., Xu, T., & Yuan, Z. (2019). Deep learning convolutional neural network for the retrieval of land surface temperature from AMSR2 data in China. *Sensors*, 19(13), 2987. <https://doi.org/10.3390/s19132987>
- Tanoori, G., Soltani, A., & Modiri, A. (2024). Predicting urban land use and mitigating land surface temperature: Exploring the role of urban configuration with convolutional neural networks. *Journal of Urban Planning and Development*, 150(3), 04024029. <https://doi.org/10.1061/JUPDDM.UPENG-5010>
- Wan, Z., & Dozier, J. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4), 892–905. <https://doi.org/10.1109/36.508406>
- Wang, J., Zhang, N., Zhang, L., Jing, H., Yan, Y., Wu, S., & Liu, R. (2024). Spatial downscaling of nighttime land surface temperature based on geographically neural network weighted regression kriging. *Remote Sensing*, 16(14), 2542. <https://doi.org/10.3390/rs16142542>
- Wang, M., Liu, W., Chu, R., Wang, X., Zhu, S., & Liao, G. (2026). HiT_DS: A modular and physics-informed hierarchical transformer framework for spatial downscaling of sea surface temperature and height. *Remote Sensing*, 18(2), 292. <https://doi.org/10.3390/rs18020292>
- Wang, Y., Huang, G., Pan, B., & others. (2024). Correcting climate model sea surface temperature simulations with generative adversarial networks: Climatology, interannual variability, and extremes. *Advances in Atmospheric Sciences*, 41(7), 111–124. <https://doi.org/10.1007/s00376-024-3288-6>
- Wei, L., Guan, L., & Zou, R. (2025). The regional sea surface temperature analysis with ultrahigh resolution produced by a deep learning-driven variational scheme. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–10. <https://doi.org/10.1109/TGRS.2025.3629740>
- Xie, B., Qi, J., Yang, S., Sun, G., Feng, Z., Yin, B., & Wang, W. (2024). Sea surface temperature and marine heat wave predictions in the South China Sea: A 3D U-Net deep learning model integrating multi-source data. *Atmosphere*, 15(1), 86. <https://doi.org/10.3390/atmos15010086>
- Xu, S., Wang, D., & Liang, S. (2024). A novel approach to estimate land surface temperature from Landsat using transfer-learning neural networks. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2024.176783>
- Xu, T., Zhou, Z., Li, Y., Wang, C., Liu, Y., & Rong, T. (2023). Short-term prediction of global sea surface temperature using deep learning networks. *Journal of Marine Science and Engineering*, 11(7), 1352. <https://doi.org/10.3390/jmse11071352>
- Ye, X., Wang, P., Zhu, J., & Yan, Y. (2025). Land surface temperature end-to-end retrieval considering the thermal environment using CNN. *IEEE Geoscience and Remote Sensing Letters*. <https://doi.org/10.1109/TGRS.2025.3525728>
- Yin, H., Zhang, D., Liu, R., & Li, H. (2025). AIGR-LSTM: Land surface temperature interpolation using adaptive inductive graph representation learning. *IEEE Geoscience and Remote Sensing Letters*, 22, Art. no. 5504905. <https://doi.org/10.1109/LGRS.2025.3560000>

- Zhang, T., Lin, P., Liu, H., Wang, P., Wang, Y., Xu, K., Zheng, W., Li, Y., Jiang, J., & Zhao, L. (2025). A coupled transformer-CNN network: Advancing sea surface temperature forecast accuracy. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 4207414. <https://doi.org/10.1109/TGRS.2025.3550827>
- Zhang, W., Jia, Z., Li, B., Liu, Q., Zhu, X., Jia, C., & Yu, T. (2024). Research on Landsat 8 land surface temperature retrieval and spatial-temporal migration capabilities based on random forest model. *Advances in Space Research*, 74(2), 610–627. <https://doi.org/10.1016/j.asr.2024.04.007>
- Zhang, X., Zhang, R., Wang, P., Dong, K., Qu, L., & Jiang, L. (2024). Estimation of 1-km all-sky land surface temperature by integration of passive microwave and thermal infrared remote sensing data through random forest. In *2024 Photonics & Electromagnetics Research Symposium (PIERS)* (pp. 1–7). IEEE. <https://doi.org/10.1109/PIERS62282.2024.10618241>
- Zhang, Z. (2026). U-MoE: A novel adaptive mixture-of-experts U-Net framework for land surface temperature reconstruction. *IEEE Geoscience and Remote Sensing Letters*. <https://doi.org/10.1109/lgrs.2026.3667345>
- Zhao, G., Song, L., Zhao, L., & Tao, S. (2024). A comparison of different machine learning methods to reconstruct daily evapotranspiration time series estimated by thermal-infrared remote sensing. *Remote Sensing*, 16(3), 509. <https://doi.org/10.3390/rs16030509>
- Zhao, H., He, X., & Li, W. (2026). Dynamic graph contrastive learning neural network for sea surface temperature prediction. *IEEE Geoscience and Remote Sensing Letters*. <https://doi.org/10.1109/LGRS.2026.3672297>
- Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., & Sobrino, J. A. (2019). Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives. *Remote Sensing*, 11(1), 48. <https://doi.org/10.3390/rs11010048>
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36. <https://doi.org/10.1109/MGRS.2017.2762307>

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.