REVIEW ARTICLE

Check for updates

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

Azad Rasul[1,2,*]

[1] Department of Geography, Faculty of Arts, Soran University, Soran 44008, Iraq
[2] Department of Forestry, College of Agricultural Engineering Sciences, Salahaddin University-Erbil, Erbil 44002, Iraq

Abstract

This systematic review, conducted in accordance with the PRISMA 2020 guidelines, maps the applications of machine learning (ML) and deep learning (DL) within thermal remote sensing (RS) from January 2018 to March 2026. Despite significant growth in this field driven by high-resolution satellite missions and open-source frameworks, no comprehensive PRISMA-compliant synthesis has previously focused exclusively on this domain during the post-2018 period. Through a structured search of Scopus and Google Scholar, we identified 193 peer-reviewed studies meeting inclusion criteria, of which 93 provided full-text access for in-depth methodological appraisal and 100 were analyzed at the metadata level. Our analysis reveals a concentrated architectural landscape, with convolutional neural networks (CNNs), long short-term memory networks (LSTM/BiLSTM), and support vector machines/regression (SVR/SVM) predominating. Application-wise, the literature skews heavily toward sea surface temperature (SST) forecasting and land surface temperature (LST) retrieval and downscaling, while critical areas such as wildfire detection, evapotranspiration estimation, and permafrost monitoring remain notably understudied. A key concern is the severe deficit in open science practices: fewer than 5% of the full-text accessible studies reported code availability, raising substantial reproducibility challenges. While the field demonstrates technical maturity, it remains architecturally conservative, presenting clear opportunities for exploring emerging approaches like physics-informed neural networks and transformer-based models. The findings underscore an urgent need for the community to address reproducibility gaps and to expand ML/DL applications into underserved domains, thereby broadening the scientific impact and operational utility of thermal RS in the coming years.

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

Submitted: 01 June 2026
Revised: 31 July 2026
Accepted: 05 August 2026
Published: 20 September 2026

Vol. 1, **No.** 2, 2026.
10.62762/JGEO.2026.270724

***Corresponding author:**
Azad Rasul
azad.rasul@soran.edu.iq

Citation
Rasul, A. (2026). Machine Learning and Deep Learning Approaches in Thermal Remote Sensing: A Systematic Review (2018–2026). *Journal of Geoscience and Earth Observation*, 1(2), 137–157.

1 Introduction

1.1 Thermal Remote Sensing: Foundations and Significance

Thermal Remote Sensing (thermal RS) 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 [1, 2]. SST plays a central role in air-sea heat exchange, ocean circulation, and the modulation of global weather and climate variability [3].

Retrieving LST and emissivity from satellite-borne TIR sensors has historically depended on physics-based algorithms. Among the most influential is the generalised split-window algorithm of Wan and Dozier [4], 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. [5] provides simultaneous retrieval of surface temperature and emissivity spectra across five TIR bands. These algorithms have delivered reliable retrievals under well-characterised conditions, but they are sensitive to assumptions about atmospheric profiles and emissivity priors – and their performance degrades under cloudy skies, complex terrain, and heterogeneous surface cover.

The satellite thermal RS landscape has shifted considerably since 2018. ECOSTRESS, launched aboard the International Space Station in June 2018, brought 70-m resolution TIR observations with irregular diurnal sampling – a capability that polar-orbiting missions simply could not offer [6]. Landsat 8 and 9 TIRS have continued delivering 100-m thermal data on a 16-day revisit cycle [7], while Sentinel-3 SLSTR and GOES-R ABI extend coverage to finer temporal resolutions. This expanding constellation has produced unprecedented volumes of TIR data, demanding analytical approaches that go well beyond what traditional physics-based methods can provide [8, 9].

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

Deep learning's rise as a transformative paradigm traces back to LeCun et al. [10] in Nature. The proliferation of open-source frameworks – TensorFlow, PyTorch – from 2015 onwards put these methods within reach of a far broader research community. In the Earth observation field, Zhu et al. [11] showed that deep learning architectures – CNNs, RNNs, and autoencoders – were consistently outperforming traditional methods across tasks as varied as scene classification, hyperspectral analysis, and change detection.

Since 2018, ML/DL methods have spread into virtually every major thermal RS application. In LST retrieval, Ye et al. [12] demonstrated end-to-end CNN retrieval directly from satellite radiances, while Jia et al. [13] coupled CNN and LSTM for spatiotemporal LST prediction over Beijing. More recently, Wang et al. [14] combined CNN and Transformer architectures in a multiscale, multidimensional collaboration network to fuse Landsat and MODIS LST products, achieving sub-1.53 K RMSE across multiple study areas, while Zhang et al. [15] embedded the physical split-window and temperature-emissivity separation formulations into a deep learning framework (DL-SW-TES), improving clear-sky LST retrieval accuracy over ECOSTRESS observations relative to the standard ECO2LTSE product. For SST forecasting, Fu et al. [16] applied deep learning to coastal temperature prediction in the South China Sea, and Song et al. [17] introduced the STVformer transformer for multivariate SST forecasting.

1.3 Research Gap and Motivation

Despite this rapid proliferation, the field still lacks a comprehensive systematic review synthesising 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 [11], spanning optical, radar, and multispectral domains without thermal-specific depth, or narrow to single applications such as urban heat islands [18, 19]. This review is distinguished from both by its cross-application, cross-architecture synthesis spanning the full thermal RS landscape – LST retrieval, downscaling, gap-filling, SST forecasting, and UHI modelling – over the post-2018 deep learning era. A rigorous, PRISMA-compliant systematic review mapping this entire landscape from 2018 to early 2026 is therefore both timely and necessary.

1.4 Objectives

Our 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 thermal RS, 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 RS.

2 Methodology

Our review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [20]. The complete search, deduplication, and screening workflow was implemented in Python using Jupyter notebooks, ensuring a fully reproducible and transparent audit trail of all decisions.

2.1 Search Strategy and Information Sources

We searched 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 chosen to align with the operational maturity of Landsat 8 TIRS, the launch of ECOSTRESS, and the widespread uptake 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 collected using the scholarly Python library with randomised inter-request delays of 8-15 seconds to respect 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 RS", "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: (1) application of at least one ML or DL method as a core analytical component – not merely referenced in passing; (2) use of thermal RS data as primary input, including satellite-derived LST, SST, TIR imagery, brightness temperature, or data from thermal-band sensors (studies integrating passive microwave data alongside TIR observations, e.g., for all-weather LST/SST estimation, were retained where TIR remained the primary retrieval mechanism, consistent with this review's focus on satellite-derived surface temperature more broadly rather than TIR exclusively); (3) publication between January 2018 and March 2026; and (4) primary research articles or systematic reviews published in peer-reviewed journals or conference proceedings.

2.2.2 *Exclusion Criteria*

A study was excluded if any one of the following conditions applied: (1) no ML/DL method employed, with reliance solely on physics-based retrieval algorithms or classical statistical models; (2) thermal data absent, i.e., studies using only visible, multispectral (non-TIR), SAR, or LiDAR data; (3) publication year outside 2018–2026; (4) not a scientific publication, including editorials, opinion pieces, or non-peer-reviewed grey literature; (5) non-Earth-observation thermal applications, such as medical thermography, industrial inspection, building energy simulation, or indoor thermal comfort; (6) thermal data derived exclusively from ground-based sensors, UAV/drone-mounted cameras, or handheld infrared cameras without satellite integration, as these platforms operate at scales, sensor specifications, and calibration standards fundamentally different from satellite-borne thermal remote sensing, precluding direct methodological comparison with the satellite-focused scope of this review; and (7) 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 pipeline implemented in Python using Jupyter notebooks – five main screening stages 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 [20]. The detailed record flow across all steps is presented in Figure 1.

2.3.1 Stage 1 – Deduplication

Records from all databases were merged into a single master dataframe. Deduplication used exact DOI matching followed by fuzzy title matching via 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

After deduplication, 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 then required 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 went through 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 co-occurring weak-domain penalty of -1. Only records in the CORE ($\geq$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, data extraction was limited to title, author-provided abstract, and structured metadata fields via the Scopus API. Critically, for 84 of these 100 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, but 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 [20].

2.5 Data Extraction Framework

Data were extracted using a structured template developed before screening began. Fields captured included six categories: (1) bibliographic information (title, authors, year, journal/conference, DOI, and source database); (2) sensor and data characteristics (satellite/sensor platform, thermal band or product, and spatial and temporal resolution); (3) ML/DL method (primary architecture, deep learning framework, and training approach); (4) application domain, comprising the primary application such as LST retrieval, LST downscaling, UHI mapping, SST forecasting, evapotranspiration, and glacier monitoring; (5) performance metrics (RMSE, MAE, R^2, and domain-specific metrics); and (6) open science practices (availability of code and data, and repository URLs).

2.6 PRISMA 2020 Record Flow

As shown in Figure 1, a total of 5,407 records were identified from Scopus and Google Scholar. After deduplication, 5,386 records remained and were sequentially screened by title, abstract, domain

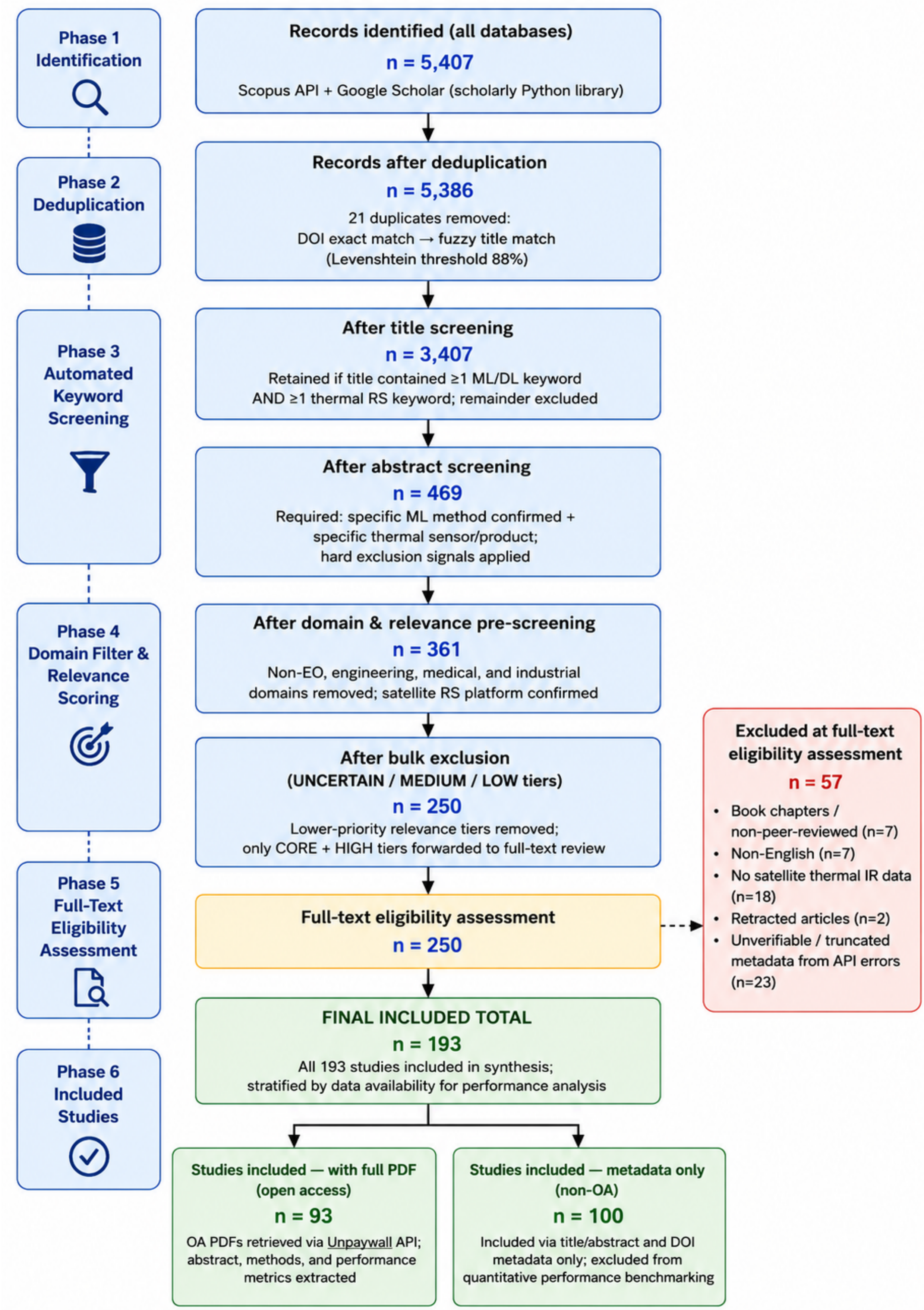


Figure 1. PRISMA 2020 flow diagram for the systematic review of ML/DL applications in thermal remote sensing (2018–March 2026).

relevance, and predefined relevance criteria. Of the 250 records forwarded to full-text eligibility assessment, 57 were excluded because they were book chapters or non-peer-reviewed publications, non-English studies, studies without satellite thermal infrared data, retracted articles, or records with unverifiable/truncated metadata. The final synthesis comprised 193 studies. Among these, 93 open-access studies had retrievable full-text PDFs and were eligible for comprehensive methodological and quantitative performance analysis, whereas 100 non-open-access studies were retained on the basis of title, abstract, and bibliographic metadata and were used primarily for descriptive analyses where sufficient information was available.

2.7 Quality Assessment and Risk of Bias

Methodological quality 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: clarity and completeness of ML/DL architecture description; independence of training and test datasets; appropriateness of performance metrics; generalisability of validation (spatial and temporal cross-validation); availability of code and data; comparison against a baseline or benchmark method; and transparency of sensor pre-processing and atmospheric correction procedures. This assessment was applied to the 93 full-text studies; the 100 metadata-only records lacked sufficient methodological detail for evaluation.

2.8 Synthesis Approach

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

3 Results

A total of 193 peer-reviewed studies published between 2018 and 2026 were included following the full PRISMA 2020 screening pipeline – 93 available as open-access full texts and 100 included on the basis of title, abstract, and structured metadata due to institutional access limitations at retrieval time. These studies span multiple thermal RS domains, principally 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 shows clear and sustained exponential growth in ML/DL applications to thermal RS over the review period (Table 1). Output was modest in the early years – just 1 study in 2018 and 5 in 2019 – reflecting the still-nascent adoption of deep learning in the thermal RS community [37, 40]. 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 sustained surge subsequently emerged in 2023 onwards: 25 studies that year, 62 in 2024, 65 in 2025, and 19 already published in the first months of 2026 – yielding a cumulative total of 193 included studies across the review period.

This acceleration reflects three converging forces: the availability of large open-access thermal satellite archives (ECOSTRESS, Landsat Collection 2, MODIS Version 6.1) [7, 8]; the maturation of PyTorch and TensorFlow ecosystems for spatiotemporal modelling; and the introduction of vision transformers and domain-adapted variants into remote sensing workflows [17, 34, 35].

3.2 Taxonomy of ML/DL Architectures

Architecture data were extractable for 177 of the 193 included studies – 93 full-text records plus 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 wider trend towards hybrid and ensemble modelling [16, 46, 72].

Architecture classifications for the 84 non-OA studies were inferred from title/abstract metadata alone and were not independently verified against full text; consequently, these classifications carry a higher risk of misclassification – particularly for studies

Table 1. 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

employing hybrid or multi-architecture approaches not fully specified in the abstract – relative to the 93 full-text-confirmed studies.

3.2.1 *Deep Convolutional Architectures*

Convolutional neural networks (CNNs) were the most frequently used architecture, appearing in 45 of the 93 full-text studies (48.4%). Their dominance is unsurprising: the strong spatial structure of thermal RS data — where local pixel-neighbourhood relationships in LST and SST fields are efficiently captured by convolutional filters — suits CNNs particularly well [21]. Standard 2D CNNs appeared across all application domains, while 3D-CNN variants were especially 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, and the novel U-MoE adaptive mixture-of-experts U-Net of Zhang et al. [22]) – appeared in 7 full-text studies. They were strongly favoured for spatially dense prediction tasks: LST downscaling, SST gap-filling, and mesoscale eddy detection. Sun and Wang [23] introduced a coordination attention residual U-Net for SST reconstruction in the South China Sea, while Qi et al. [24] proposed the RVDCU-Net framework combining residual and dilated convolution modules for spatiotemporal SST reconstruction.

Generative adversarial networks (GANs) appeared in 11 full-text studies (11.8%), used mainly for thermal super-resolution and SST downscaling. Wang et al. [25] used CycleGAN to correct climate model SST simulations, improving climatological biases, interannual variability including ENSO, and SST extremes. Jalbuena et al. [26] applied a Regression-based GAN (RGAN) to correct landcover bias in MODIS LST downscaling, outperforming traditional methods across urban areas in Korean cities. Sundar et al. [27] developed a hybrid GAN-RNN model integrating remote sensing data for urban environmental analysis and resilience. GANs are among the fastest-growing methods – absent before 2022, with publications nearly doubling 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. They proved well-suited to cloud-gap reconstruction, where the objective is to infer missing thermal values from partial spatial observations [28].

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 (36.6%), of which 7 employed ConvLSTM. LSTM models were most prevalent in SST forecasting tasks, where temporal autocorrelation at daily-to-monthly timescales could be exploited [29]. Li et al. [30] used LSTM to predict urban LST under land use change scenarios, while Yin et al. [31] proposed AIGR-LSTM for LST interpolation combining adaptive information granulation with recurrent learning. El Azhary and Minaoui [32] 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, 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 family, appearing in 13 full-text studies. Their adoption followed a clear chronological trajectory: absent before 2022, first appearing in SST forecasting in 2023, then expanding into LST retrieval and downscaling by 2024-2026. Jia et al. [33] introduced TL-iTransformer for SST prediction with transfer learning, while Song et al. [17] proposed the STVformer spatial-temporal-variable transformer for improved SST forecasting. Hu et al. [34] proposed THSTNet – a two-stage hierarchical spatiotemporal fusion network using the Swin Transformer – to fuse MODIS and Landsat LST products for fine-resolution (30-m) LST reconstruction, achieving 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. [35] 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 and a physics-constrained branch grounded in the mixed-layer heat budget equation, significantly outperforming ConvLSTM, DeepONet, and FNO baselines.

3.2.4 Classical Supervised Methods

Despite deep learning's dominance, classical supervised ML methods remained important, particularly in LST-related applications where interpretability is prioritised [36]. Random Forest (RF) appeared in 15 full-text studies, predominantly for LST downscaling and UHI-LST relationship modelling. Support Vector Machine/Regression (SVM/SVR) appeared in 30 studies, frequently as a baseline comparator against deep learning; SVR consistently underperformed CNN and LSTM models in spatiotemporally complex tasks but remained competitive in data-scarce settings. Artificial Neural Networks (ANN/MLP) appeared in 17 studies across all application domains, either as baseline models or as the final regression head within hybrid architectures [37].

3.2.5 Physics-Informed and Hybrid Approaches

Physics-guided neural networks appeared in several full-text studies. Dong et al. [38] 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 degrees C for daily and 0.60 degrees C for monthly forecasts. Shi et al. [39] developed PANN, a physics-guided attention-based neural network for SST prediction in the East China Sea, embedding PDE-based physical constraints via a dedicated physical constraint module and fusing them with data-driven predictions through an attention fusion module. Transfer learning was employed in 7 full-text studies (Table 2); Xu et al. [7] demonstrated a TL approach leveraging radiative transfer model simulations to improve LST retrieval from Landsat TOA data, reducing dependence on land surface emissivity inputs and outperforming conventional ML, single-channel, and split-window algorithms.

3.3 Application Domain Coverage

Application domain classification was possible for 127 of the 193 included studies; the remaining 66 lacked sufficient metadata for reliable domain assignment (see Table 3). The distribution shows 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 LST Retrieval from Satellite Thermal Sensors ($n \approx 46$, $\sim$36%)

LST retrieval was the largest single application domain (Table 3). LST retrieval studies applied ML/DL to derive land surface temperature directly from satellite-measured top-of-atmosphere radiances or brightness temperatures (Table 3). CNNs dominated here. Ye et al. [12] demonstrated end-to-end CNN-based LST retrieval that accounts for the thermal environment in sensor calibration, while Zhang et al. [41] integrated deep learning with split-window and TES algorithms for LST retrieval from Landsat 8. A notable trend was multi-source sensor fusion: 11 of 23 full-text retrieval studies combined data from two or more thermal sensors [42]. Beyond retrieval itself, LST products derived from ML/DL models have also been applied to downstream environmental modelling, such as the high-resolution mapping of ground-level ozone using interpretable machine learning approaches [43].

3.3.2 Sea Surface Temperature Forecasting and Prediction ($n \approx 29$, $\sim$23%)

Studies ranged from short-term (1–7 day) coastal forecasts – where LSTM and ConvLSTM models

Table 2. 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	48.4%	LST retrieval, SST gap-filling, downscaling
LSTM / BiLSTM / ConvLSTM	34	36.6%	SST forecasting, LST time-series prediction
Support Vector Machine / SVR	30	32.3%	Baseline comparator, SST prediction, LST retrieval
Artificial Neural Network (ANN/MLP)	17	18.3%	All domains — baseline and regression head
Random Forest (RF)	15	16.1%	LST downscaling, UHI modelling
Generative Adversarial Network (GAN)	11	11.8%	SST / LST super-resolution, downscaling
Transformer / Vision Transformer (ViT)	13	14.0%	SST forecasting, eddy detection, LST retrieval
Gradient Boosting (XGBoost/LightGBM)	9	9.7%	UHI, urban morphology–LST analysis, SST retrieval
Encoder-Decoder / U-Net	7	7.5%	LST downscaling, SST reconstruction
Transfer Learning	7	7.5%	ECOSTRESS, Sentinel-3, cross-sensor adaptation
Autoencoder / DINCAE	4	4.3%	SST / LST gap-filling
Physics-Informed NN (PINN)	3	3.2%	LST retrieval, SST prediction

Table 3. Distribution of included studies by application domain across all 193 studies. Studies without sufficient metadata for domain classification are listed as not categorized. Percentages are approximate given partial metadata availability.

Application Domain	n (all 127)	%	Dominant Architecture	Key Sensors
LST Retrieval + Related	46	36.2%	CNN, Transformer, Hybrid	MODIS, ASTER, ECOSTRESS, Landsat TIRS
SST Forecasting / Prediction	29	22.8%	ConvLSTM, BiLSTM, Transformer, GNN	NOAA AVHRR, GOES, AMSR
LST Downscaling / Fusion	23	18.1%	CNN, GAN, Random Forest	MODIS, Landsat, Sentinel-3
LST / SST Gap-Filling	20	15.7%	CNN, Autoencoder, ConvLSTM	MODIS, Himawari AHI, AMSR
Urban Heat Island (UHI)	5	3.9%	Random Forest, XGBoost, LSTM, Xception	Landsat 8/9 TIRS, MODIS, ECOSTRESS
Evapotranspiration / Agriculture	2	1.6%	Deep NN, LSTM	ECOSTRESS, Landsat, MODIS
Permafrost / Glacier Temperature	2	1.6%	SVM, ANN	Landsat TIRS, ASTER
Total classified	127	100%	—	—

Note: Of the 193 included studies, 66 (34.2%) lacked sufficient metadata for domain classification and are excluded from this table. Percentages above are calculated using the 127 classifiable studies as the denominator, not the full sample of 193, to avoid underestimating individual domains' relative representation.

typically achieved RMSE values of 0.4–0.9 degrees C in marginal and semi-enclosed seas – to medium-range monthly predictions, such as the EMD-LSTM model of Cai et al. [29], which reported RMSE within 0.5 degrees C for monthly average SST in the South China Sea. Focusing on daily timescales in the same region, Qi et al. [44] applied a deep learning-driven approach to forecast daily SST, further demonstrating the applicability of neural networks to high-frequency thermal prediction in marginal seas. At the longer end of the spectrum, Krestenitis et al. [45] introduced a deep learning framework for interim-future (1-year-ahead) daily SST forecasting over the northeastern Mediterranean, trained on 15 years of satellite data (2008-2022) and demonstrating skill even for data more than two years beyond the training period.

A recurring finding across SST forecasting studies was that spatiotemporal hybrid architectures consistently outperformed purely spatial or temporal models. Zhang et al. [46] introduced a coupled Transformer-CNN network for SST forecasting, while Fu et al. [16] combined LSTM with Transformer attention layers to improve SST prediction accuracy in regional seas. Khalilabadi and Sadati [47] further developed a new hybrid deep learning-based method for SST prediction, demonstrating the continued evolution of ensemble approaches in this domain. Jin et al. [48] 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. [49] extended graph-based approaches further with Dynamic Graph Contrastive Learning for SST forecasting.

3.3.3 LST Downscaling and Spatiotemporal Fusion ($n \approx$ 23, $\sim$18%)

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. [50] explored scale-dependency effects in LST downscaling using a CNN-ELM hybrid, and Rogalski and Ilunga [51] showed that Random Forest can predict LST from rainfall and elevation across South African dryland ecosystems. Gradient boosting methods have also proven effective; for instance, Khanifar and Khademalrasoul [52] applied the XGBoost algorithm to model the relationship between LST and multiscale curvatures in southwestern Iran, demonstrating the utility of tree-based ensembles for capturing complex topographic controls on surface temperature. GAN-based super-resolution approaches [26] consistently produced sharper thermal boundaries than CNN regression approaches, though at higher computational cost. Beyond GANs, alternative deep learning-driven schemes have also been developed for thermal field downscaling; for instance, Wei et al. [53] proposed a deep learning-driven variational scheme for regional SST analysis with ultrahigh resolution, demonstrating the potential of integrating learning-based priors with variational frameworks.

3.3.4 LST and SST Gap-Filling and Cloud Reconstruction ($n \approx 20$, $\sim$16%)

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 while better preserving spatiotemporal variability. El Azhary and Minaoui [32] proposed EDDA-ConvLSTM for mesoscale SST reconstruction in the Mediterranean with dual-attention mechanisms. Han et al. [42] developed a time-continuous LST data fusion method for MODIS-Landsat spatiotemporal reconstruction that reduces cloud-induced gaps in urban monitoring. Zhang et al. [54] proposed a random forest-based approach integrating passive microwave and thermal infrared data to estimate all-sky 1-km LST, effectively addressing cloud contamination through multi-source sensor fusion. Ge et al. [55] predicted next-day 1-km canopy urban heat island maps using CNN-based spatiotemporal fusion, showing that near-real-time thermal urban monitoring is feasible.

3.3.5 Urban Heat Island (UHI) Modelling ($n \approx 5$, $\sim$4%)

UHI studies applied ML/DL to characterise the spatial drivers of surface urban heat islands and forecast future UHI intensity. Singh et al. [56] introduced the hybrid Xception-LSTM model for advanced urban heat island detection from satellite thermal imagery. Lin [57] applied deep learning to integrate remote sensing data for UHI distribution observation in Chinese cities. Ge et al. [55] 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. [58] proposed a spatiotemporal deep learning approach to enhance air temperature estimation using only land surface temperature as input. In addition to deep learning, classical ML methods remain valuable for UHI analysis; Li [59] compared spatial regression models and random forest for evaluating the correlation between impacting factors and LST, providing insights into the relative strengths of parametric and non-parametric approaches for urban thermal environment studies. Lu et al. [60] 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. Similarly, Chen et al. [61] evaluated the seasonal effects of

Table 4. Thermal satellite sensors and platforms used across the 54 reviewed studies. Because many studies employed multiple sensors, percentages exceed 100%.

Standardized thermal sensor/platform	**Number of studies (n)**	**% of 54 studies**	**Typical application**
Landsat (TM/ETM+/TIRS)	25	46.3%	LST retrieval, downscaling, UHI analysis
MODIS (Terra/Aqua)	23	42.6%	LST retrieval, SST, downscaling, gap filling
Passive microwave LST/SST	—	—	—
AMSR-E/AMSR2	6	11.1%	Passive microwave LST/SST retrieval
VIIRS	4	7.4%	SST/LST retrieval and super-resolution
AVHRR	4	7.4%	SST retrieval and reconstruction
ASTER	3	5.6%	High-resolution LST retrieval and validation
ECOSTRESS	3	5.6%	High-resolution LST estimation
Meteosat SEVIRI	3	5.6%	Geostationary LST/SST
Sentinel-3 SLSTR	2	3.7%	SST/LST retrieval
Himawari AHI	1	1.9%	SST retrieval
GOES-R ABI	1	1.9%	Near-surface temperature estimation
FY-2G	1	1.9%	Geostationary LST
FY-3 EMERSI-LL	1	1.9%	All-sky LST estimation
AATSR	1	1.9%	SST retrieval
ATMS	1	1.9%	All-weather SST retrieval
AGRI	1	1.9%	LST retrieval

building form and street view indicators on street-level LST using random forest regression, highlighting the importance of fine-scale urban morphology in modulating thermal conditions.

3.3.6 *Evapotranspiration, Agriculture, and Other Domains ($n \approx 4$, ∼3%)*

Four full-text studies addressed thermal RS applications outside the core LST/SST domains. Jawad et al. [62] developed a hybrid Penman-Monteith and deep learning (PMDL) framework for evapotranspiration estimation over drylands, achieving high accuracy ($R^2 > 0.85$; KGEss > 0.78) and outperforming widely used products such as MOD16A2, ECO3ETPTJPL, ECO3ETALEXI, and OpenET. Zhao et al. [63] assessed four ML models – deep forest (DF), deep neural networks (DNN), random forest (RF), and extreme gradient boosting (XGB) – for reconstructing ET products and filling gaps over the Heihe River Basin. RF showed the highest robustness (R = 0.73 with ground observations); DNN and XGB also performed well (R > 0.70), though residual gaps persisted in desert regions for XGB, and DF could reconstruct the full basin but tended to underestimate ET (R = 0.66). Karimian et al. [64] monitored spatiotemporal land use-land cover changes and their relationship to LST using deep learning in semiarid environments. Li et al. [65] simulated the LST impacts of proposed land use and land cover plans using an integrated deep neural network approach, demonstrating the potential of DL for scenario-based thermal environment forecasting in urban energy planning. Glacier and permafrost temperature monitoring studies applied SVM and ANN to Landsat TIRS and ASTER data over Southeast Asian highland regions.

The "Not categorised" group (studies lacking sufficient title/abstract metadata for domain classification, largely among the 100 non-open-access records) does not appear concentrated in any single domain and is therefore unlikely to materially bias the observed application-domain imbalance.

3.4 Sensor and Platform Usage

Thermal satellite sensor usage was identified from the 54 studies that explicitly employed satellite thermal sensors. As shown in Table 4, Landsat (TM/ETM+/TIRS) and MODIS (Terra/Aqua) were the dominant platforms, appearing in 46.3% and 42.6% of the reviewed studies, respectively. Their widespread adoption reflects their long-term data continuity, global coverage, and established suitability for land surface temperature retrieval, downscaling, urban heat island analysis, and related deep learning applications.

Passive microwave sensors, including AMSR-E/AMSR2, were used in 11.1% of the studies, primarily for all-weather land and sea surface temperature retrieval. Other thermal sensors, such as VIIRS, AVHRR, ECOSTRESS, Meteosat SEVIRI, and Sentinel-3 SLSTR, were employed less frequently but provided important capabilities for applications including high-resolution temperature estimation, geostationary monitoring, and sea surface temperature reconstruction. Sensors such as Himawari AHI, GOES-R ABI, FY-2G, FY-3E MERSI-LL, AATSR, ATMS, and AGRI appeared only occasionally, reflecting their more specialized application domains or more limited use within the reviewed literature.

Notably, after manually verifying the reviewed papers, ASTER was identified in only three studies (5.6%), rather than serving as the dominant platform. This finding indicates that ASTER plays a comparatively limited role in recent deep learning-based thermal remote sensing studies. Furthermore, studies relying exclusively on reanalysis products, merged SST datasets, climate model outputs, or other non-sensor datasets were excluded from this sensor-use analysis to ensure that Table 4 reflects only actual satellite thermal sensors.

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. These RMSE, MAE, and R^2 values should be interpreted cautiously, as they are not directly comparable across studies given differing tasks, sensors, spatial resolutions, forecast horizons, and validation references.

3.6 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 (Table 5). 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 [69] 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 [70]; 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.

3.7 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 – targeting spatiotemporally continuous LST reconstruction rather than spatial downscaling per se – 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 [71]. Liang et al. [67] 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. [68] 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. [66] 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.

3.8 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. [73] 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. [74] 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

Table 5. 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. [68]; He et al. [81]
LST Retrieval	R^2	0.99	0.69–0.99	Zhang et al. [41]; Xu et al. [7]
LST Downscaling	RMSE (K)	0.886	0.886–3.23	Wang et al. [66]; Fu et al. [71]
LST Downscaling	R^2	0.97	0.58–0.97	Liang et al. [67]; Fu et al. [71]
SST Forecasting	RMSE (K)	0.084	0.084–1.40	Xu et al. [73]; Jiao et al. [74]
SST Forecasting	R^2	0.996	0.956–0.996	Shi et al. [83]; Rani et al. [72]
SST Gap-Filling	RMSE (K)	0.463	0.46–2.07	Barth et al. [28]; Fan et al. [77]
SST Gap-Filling	MAE (K)	0.023	0.023–1.22	Goh et al. [82]; Fan et al. [77]
UHI / LST Mapping	RMSE (K)	1.042	1.042–1.65	Singh et al. [56]; Kporha et al. [80]
UHI / LST Mapping	R^2	0.966	0.966	Ge et al. [55]; Lu et al. [60]

error compared to numerical models. Xie et al. [75] 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. [76] 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.

3.8.1 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 [77]. Putra and Hsu [78] 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 = 0.463 K, handling variable-accuracy missing data through likelihood-maximisation training [28].

3.9 Methodological Quality and Open Science Practices

Methodological reporting quality varied considerably across the 93 full-text studies. Common limitations included: single-site or single-region validation without spatial cross-validation, present in approximately 60% of LST downscaling and UHI studies; absence of baseline comparisons against physics-based algorithms in 31% of LST retrieval studies; non-standardised performance metrics that complicate cross-study comparison; and data leakage risk in spatiotemporal settings where temporal overlap between training and test sets was not always explicitly excluded – one LST retrieval study applied a random pixel-level division of the full time series without chronological separation [79], 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 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, 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 careful discussion. The following subsections address each in turn, contextualising findings within the broader ML/DL remote sensing literature [10, 11].

4.1 Architectural Conservatism and the Transformer Transition

The dominance of CNN and LSTM architectures – present in 48.4% and 36.6% of full-text included studies, respectively – reflects the maturity of these frameworks for spatiotemporal thermal RS modelling, but arguably also indicates a degree of conservatism in the community's adoption of newer approaches. 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 [70]), remained a minority of the corpus (13 full-text studies, 14%). That gap is likely to close rapidly: the inclusion of five transformer-based studies published in 2025-2026 alone [14, 83–86] suggests an accelerating uptake trajectory consistent with the broader remote sensing community's shift towards transformer architectures.

Evidence of frontier architectures was already detectable in the 2025-2026 literature. The HiT_DS hierarchical Transformer of Wang et al. [86] 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 yet represented in the confirmed corpus, and formal benchmark evaluations for transformer architectures on thermal RS tasks remain sparse. The scarcity of foundation models in thermal RS likely stems from a domain gap – most are pre-trained on optical imagery with different radiometric properties – compounded by the lack of large, standardised thermal datasets needed to pre-train thermal-specific alternatives.

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) [3], while wildfire detection and permafrost monitoring require event-triggered or site-specific data collection.

The UHI domain – well-represented in traditional remote sensing literature [18, 19] – accounted for only approximately 4% of the 127 classifiable studies. This 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 – a surprisingly low figure given compelling theoretical arguments for their advantages in thermal RS. The radiative transfer equation governing satellite LST retrieval is well-characterised [4, 5]. 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 [62], and ocean heat transport dynamics constrain SST variability. The MDK-DL framework demonstrated that physics embedding can improve cross-sensor generalisation in LST retrieval [69]. Dong et al. [38] and Shi et al. [39] provide compelling recent examples of physics-guided DL for SST forecasting and prediction, suggesting the barrier to PINN adoption is beginning to lower.

Beyond retrieval, PINNs hold particular promise for tasks constrained by well-established physical relationships. In LST downscaling, embedding energy balance or radiative transfer constraints could reduce the reliance on purely statistical scale-dependent relationships that often fail to generalise across landscapes. Similarly, all-weather LST reconstruction under cloud cover could benefit from physical priors linking LST to net radiation and atmospheric state, offering a principled alternative to purely data-driven gap-filling. For evapotranspiration estimation, coupling PINNs with the Penman-Monteith or surface energy balance framework could improve physical consistency and cross-site transferability relative to black-box regression approaches.

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, results cannot be independently verified – and this is particularly acute for operational applications such as LST gap-filling [32, 42] and SST forecasting [17, 45], where reproducibility of model training pipelines is a prerequisite for operational deployment. The persistently low code availability across the corpus suggests that voluntary adoption of open science practices is not enough; journal-level enforcement of data and code sharing policies may be necessary.

4.5 Implications of the Data Access Stratification

The stratified inclusion of 93 full-text and 100 metadata-only studies is a deliberate methodological transparency choice, consistent with PRISMA 2020 best practices [20]. The 100 non-OA studies are not absent from our review – they appear 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 bear in mind 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.

As a sensitivity check, restricting the architecture distribution to only the 93 full-text studies yields substantially similar patterns to the full 193-study sample (CNN and LSTM/BiLSTM remain the dominant architectures). Restricting the application-domain distribution to only the 93 full-text studies likewise yields a similar pattern to the 127-study classifiable sample (LST retrieval and SST forecasting remain the two largest application domains), suggesting that title/abstract-based classification of the metadata-only studies did not materially distort the review's main conclusions.

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 Remote Sensing

Hu et al. [34] and Guo et al. [35] have demonstrated the potential of transformer-based architectures for thermal spatiotemporal fusion and SST forecasting, but systematically 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. [38], and Shi et al. [39] 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 [4] and TES [5] algorithms. Developing 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 [62], 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 architectural comparison across studies and makes it difficult to identify 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 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 [21, 51, 64]. Systematic evaluation of model transferability across climate zones, biomes, and sensor systems is needed before operational deployment can be justified. Adopting 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

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. It should be noted that 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 ML/DL to thermal RS 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 [6, 7], 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 (CNN: 48.4%; LSTM/BiLSTM/ConvLSTM: 36.6% of full-text included studies), with SVM/SVR as the third most common method (32.3%). 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 (˜36%), SST forecasting (˜23%), LST downscaling (˜18%), and LST/SST gap-filling (˜16%) of the 127 domain-classifiable studies. Wildfire detection, evapotranspiration estimation [62], permafrost monitoring, and glacier thermal dynamics – collectively accounting for approximately 3% of classifiable studies – remain 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. [38] achieved RMSE of 0.17 degrees C for daily global SST forecasts; Shi et al. [39] 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. Broader PINN adoption – facilitated by shared, differentiable implementations of the radiative transfer equation – represents the single most impactful near-term methodological opportunity identified in this review.
- Open science practices were critically underadopted: code availability was explicitly stated in fewer than 5% of full-text 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 [20]. Future systematic updates conducted with broader institutional access would strengthen the completeness of the performance benchmarking synthesis.

Taken together, 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 through standardised 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. At the same time, 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.

Data Availability Statement

Not applicable.

Funding

This work was supported without any funding.

Conflicts of Interest

The author declares no conflicts of interest.

AI Use Statement

The author declares that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] 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. [CrossRef]

[2] Norman, J. M., & Becker, F. (1995). Terminology in thermal infrared remote sensing of natural surfaces. *Agricultural and Forest Meteorology*, *77*(3-4), 153-166. [CrossRef]

[3] 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. [CrossRef]

[4] 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. [CrossRef]

[5] Gillespie, A., Rokugawa, S., Matsunaga, T., Cothern, 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. [CrossRef]

[6] Fisher, J. B., Lee, B., Purdy, A. J., Halverson, G. H., Dohlen, M. B., Cawse-Nicholson, K., ... & Hook, S. (2020). ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the international space station. *Water Resources Research*, *56*(4), e2019WR026058. [CrossRef]

[7] Xu, S., Wang, D., Liang, S., Jia, A., Li, R., Wang, Z., & Liu, Y. (2024). A novel approach to estimate land surface temperature from Landsat top-of-atmosphere reflective and emissive data using transfer-learning neural network. *Science of the Total Environment*, *955*, 176783. [CrossRef]

[8] 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. [CrossRef]

[9] He, X., Liu, X., Ru, C., Deng, X., Zhao, R., & Yu, W. (2025). Evaluating spatial representativeness across multiple scales for a comprehensive ground validation network using Landsat land surface temperature data and random forest. *IEEE Transactions on Geoscience and Remote Sensing*, *63*, 1-14. [CrossRef]

[10] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. [CrossRef]

[11] 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. [CrossRef]

[12] Ye, X., Wang, P., Zhu, J., Duan, Y., & Yang, B. (2025). Land Surface Temperature end-to-end retrieval considering the topographic effect using radiative transfer model-driven convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *63*, 1-10. [CrossRef]

[13] Jia, H., Yang, D., Deng, W., Wei, Q., & Jiang, W. (2021). Predicting land surface temperature with geographically weighed regression and deep learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(1), e1396. [CrossRef]

[14] Wang, M., Chen, X., Wang, G., Wang, R., Du, L., & Zhang, Z. (2026). MSDCN-LSTF: A Multiscale and Multidimensional Collaboration Network for Land Surface Temperature Spatiotemporal Fusion Based on Transformer and Convolution. *IEEE Transactions on Geoscience and Remote Sensing*, *64*, 1-22. [CrossRef]

[15] Zhang, H., Hu, T., Tang, B. H., Mallick, K., Zheng, X., Wang, M., ... & Sobrino, J. A. (2025). Deep learning coupled with split window and temperature-emissivity separation (DL-SW-TES) method improves clear-sky high-resolution land surface temperature estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *225*, 1-18. [CrossRef]

[16] 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. [CrossRef]

[17] Song, D., Dai, S., Li, W., Ren, T., Wei, Z., & Liu, A. A. (2024). STVformer: A spatial-temporal-variable transformer with auxiliary knowledge for sea surface temperature prediction. *Applied Ocean Research*, *153*, 104218. [CrossRef]

[18] 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. [CrossRef]

[19] Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., ... & Sobrino, J. A. (2018). Satellite remote sensing of surface urban heat islands: Progress, challenges, and perspectives. *Remote Sensing*, *11*(1), 48. [CrossRef]

[20] 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*. [CrossRef]

[21] 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. [CrossRef]

[22] Zhang, Z., Mao, K., Zhai, S., & Shi, J. (2026). U-MoE: A Novel Adaptive Mixture-of-Experts U-Net Framework for Land Surface Temperature Reconstruction. *IEEE Geoscience and Remote Sensing Letters*, *23*, 1-5. [CrossRef]

[23] 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. [CrossRef]

[24] Qi, T., Tang, S., & Ye, H. (2026). RVDCU-Net: A Modular Deep Learning Framework for Spatiotemporal Reconstruction of Satellite-Derived Sea Surface Temperature. *Marine Geodesy*, *49*(3), 439-464. [CrossRef]

[25] Wang, Y., Huang, G., Pan, B., Lin, P., Boers, N., Tao, W., ... & Li, H. (2024). Correcting climate model sea surface temperature simulations with generative adversarial networks: Climatology, interannual variability, and extremes. *Advances in Atmospheric Sciences*, *41*(7), 1299-1312. [CrossRef]

[26] 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. [CrossRef]

[27] Sundar, G., Patchaiammal, P., Lakkshmanan, A., Ramya, S., Pallavi, L., & Thiagarajan, R. (2024). Comprehensive earth system analysis for advancing community resilience: integrating remote sensing with hybrid GAN-RNN techniques. *Remote Sensing in Earth Systems Sciences*, *7*(4), 788-803. [CrossRef]

[28] 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. [CrossRef]

[29] Cai, Z., Ren, K., Li, B., Guo, W., Hui, G., & Zhang, X. (2025, June). 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. [CrossRef]

[30] 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. [CrossRef]

[31] 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*, 1-5. [CrossRef]

[32] 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*, *22*, 1-5. [CrossRef]

[33] Jia, W., Guan, S., & Xue, Y. (2024). TL-iTransformer: Revolutionizing sea surface temperature prediction through iTransformer and transfer learning. *Earth Science Informatics*, *17*(5), 4847-4857. [CrossRef]

[34] 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*, *63*, 1-20. [CrossRef]

[35] 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. [CrossRef]

[36] 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*(3), 328. [CrossRef]

[37] Shatnawi, N., & Abu Qdais, H. (2019). Mapping urban land surface temperature using remote sensing techniques and artificial neural network modelling. *International Journal of Remote Sensing*, *40*(10), 3968-3983. [CrossRef]

[38] 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. [CrossRef]

[39] Shi, B., Feng, L., He, H., Hao, Y., Peng, Y., Liu, M., ... & Liu, J. (2024). A physics-guided attention-based neural network for sea surface temperature prediction. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1-13. [CrossRef]

[40] 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. [CrossRef]

[41] 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. [CrossRef]

[42] 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. [CrossRef]

[43] 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. [CrossRef]

[44] 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. [CrossRef]

[45] 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*(2), 149-168. [CrossRef]

[46] Zhang, T., Lin, P., Liu, H., Wang, P., Wang, Y., Xu, K., ... & Zhao, L. (2025). A coupled transformer-CNN network: Advancing sea surface temperature forecast accuracy. *IEEE Transactions on Geoscience and Remote Sensing, 63*, 4207414. [CrossRef]

[47] Khalilabadi, M. R., & Sadati, S. (2025). A new hybrid deep learning-based method for sea surface temperature prediction. *Ships and Offshore Structures*, 1-8. [CrossRef]

[48] 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, 22*, 1-5. [CrossRef]

[49] Zhao, H., He, X., & Li, W. (2026). Dynamic Graph Contrastive Learning Neural Network for Sea Surface Temperature Prediction. *IEEE Geoscience and Remote Sensing Letters, 23*, 1-5. [CrossRef]

[50] 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, 17*(2), 1093-1104. [CrossRef]

[51] 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). [CrossRef]

[52] 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*(12), 11763-11774. [CrossRef]

[53] 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. [CrossRef]

[54] Zhang, X., Zhang, R., Wang, P., Dong, K., Qu, L., & Jiang, L. (2024, April). 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. [CrossRef]

[55] 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. [CrossRef]

[56] Singh, A. K., Deepika, C. L., Shahnaz, K. V., Bhagyalakshmi, L., Sharada, K., Sarupriya, S., & Suman, S. K. (2025). Hybrid Xception-LSTM Model for Remote Sensing: Advanced Urban Heat Island and Land Use Analysis. *Remote Sensing in Earth Systems Sciences, 8*(1), 132-144. [CrossRef]

[57] 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), 1-17. [CrossRef]

[58] 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. [CrossRef]

[59] 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. [CrossRef]

[60] 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. [CrossRef]

[61] 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, 245*, 110884. [CrossRef]

[62] 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. [CrossRef]

[63] 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. [CrossRef]

[64] Karimian, R., Rangzan, K., Karimi, D., & Einali, G. (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*(11), 2461-2481. [CrossRef]

[65] 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. [CrossRef]

[66] 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. [CrossRef]

[67] 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. [CrossRef]

[68] Khedher, I., Favreau, J. M., Miguet, S., & Gesquière, G. (2023, December). A multimodal deep learning approach for high-resolution Land Surface Temperature estimation. In *International Conference on Applied Soft Computing and Communication Networks* (pp. 363-377). Singapore: Springer Nature Singapore. [CrossRef]

[69] Wang, H., Mao, K., Yuan, Z., Shi, J., Cao, M., Qin, Z., ... & Tang, B. (2021). A method for land surface temperature retrieval based on model-data-knowledge-driven and deep learning. *Remote sensing of environment, 265*, 112665. [CrossRef]

[70] 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. [CrossRef]

[71] 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. [CrossRef]

[72] Rani, B. S., Narayanan, D. S., Jasmine, M. 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*, 101632. [CrossRef]

[73] 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. [CrossRef]

[74] 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. [CrossRef]

[75] 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. [CrossRef]

[76] 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. [CrossRef]

[77] 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. [CrossRef]

[78] Putra, D. P., & Hsu, P. C. (2024). Leveraging Transfer Learning and U-Nets Method for Improved Gap Filling in Himawari Sea Surface Temperature Data Adjacent to Taiwan. *ISPRS International Journal of Geo-Information, 13*(5), 162. [CrossRef]

[79] 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. [CrossRef]

[80] 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. [CrossRef]

[81] He, T., Liu, F., Wang, A., & Fei, Z. (2023). Estimating monthly surface air temperature using MODIS LST data and an artificial neural network in the Loess Plateau, China. *Chinese Geographical Science, 33*(4), 751-763. [CrossRef]

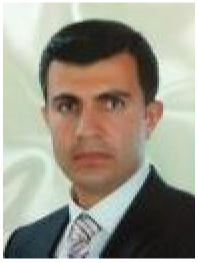
[82] Goh, E., Yepremyan, A., Wang, J., & Wilson, B. (2024). MAESSTRO: masked autoencoders for sea surface temperature reconstruction under occlusion. *Ocean Science, 20*(5), 1309-1323. [CrossRef]

[83] 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, 19*, 7714-7725. [CrossRef]

[84] 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, 18*, 15249-15264. [CrossRef]

[85] 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, 13*, 143509-143521. [CrossRef]

[86] 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. [CrossRef]



Azad Rasul is a Senior Assistant Professor of Remote Sensing at Soran University and Salahaddin University-Erbil, Iraq, with over 18 years of experience. He holds a PhD from the University of Leicester (2016) and has authored 80+ publications with over 12,000 citations in journals including The Lancet. He is a GBD Collaborator at the University of Washington and has taught over 103,000 students worldwide through online courses. (Email: azad.rasul@soran.edu.iq)