Journal of Metals, Materials and Minerals, 37(1), e2693, 2027

Intelligent material recognition from thermal signatures using hybrid deep learning frameworks

Venkateswarlu SUNKARI[1], T Kishore BABU[2], V. N. S. R. MURTHY[3], Boddu L. V. Siva Rama KRISHNA[4], P. V. V. S. D. NAGENDRUDU[5], and Sriharsha VIKRUTHI[6,*]

[1] *Department of Electrical and Computer Engineering, College of Engineering and Architecture, University of Nizwa, Birkat Al Mouz, Nizwa 616, Oman*
[2] *Department of AI&DS, KLEF (Koneru Lakshmaiah Education and Foundation), Vaddeswram, Guntur 522302, India*
[3] *Department of CSE-IoT, Ramachandra College of Engineering (A), Eluru, Andhra Pradesh 534007, India*
[4] *Department of Computer Science and Engineering, SRM University-AP, Amaravati, Andhra Pradesh, 522240, India*
[5] *Department of Artificial Intelligence and Machine Learning, Aditya University, Surampalem, Andhra Pradesh 533437, India*
[6] *Department of CSE, B V Raju Institute of Technology, Narsapur, Medak, Telangana 502313, India*

***Corresponding author e-mail**: sriharsha.vikruthi@gmail.com

Received date:
28 March 2026
Revised date:
4 April 2026
Accepted date:
23 April 2026

Keywords:
Long-wave infra-red;
Material recognition;
Deep learning architecture;
Neural networks

Abstract

Accurate material identification using thermal infrared signatures has significant implications for industrial inspection, autonomous systems, and non-contact diagnostics. In this study, we propose a physics-informed deep learning framework for intelligent material recognition based on transient thermal behavior. A large-scale dataset, ThermalMat-85K, comprising 85,000 long-wave infrared (LWIR) images across ten material classes, was constructed under controlled multi-angle, multi-distance, and multi-environment conditions. Each sample was validated through Fourier transform infrared (FT-IR) spectroscopy and X-ray diffraction (XRD) to ensure chemical and structural integrity prior to thermal acquisition. Seven deep learning architectures were systematically evaluated, including VGG-16, ResNet-50, MobileNet-V3, DenseNet-121, EfficientNet-B4, a CNN+LSTM temporal hybrid, and a proposed Hybrid Net integrating convolutional, transformer, and recurrent modules with cross-modal attention gating. EfficientNet-B4 achieved the highest Top-1 accuracy (93.1%) and macro F1-score (92.8%), demonstrating superior class-wise stability. Residual and dense connectivity architectures also exhibited strong performance, while lightweight and temporal hybrids revealed trade-offs between efficiency and discriminative power. Class-wise confusion analysis indicates that misclassification trends correlate strongly with thermophysical proximity, particularly thermal conductivity and emissivity overlap. These findings confirm that transient thermal decay dynamics encode physically meaningful features suitable for material discrimination. The results establish a robust benchmark for physics-informed thermal material recognition and provide guidance for selecting architectures based on deployment constraints and robustness requirements.

1. Introduction

Material identification plays a fundamental role in industrial quality control, structural health monitoring, automated sorting systems, robotics, and defense applications [1,2]. Accurate recognition of material type is essential for ensuring manufacturing consistency, detecting structural degradation, enabling autonomous manipulation, and supporting non-contact inspection in hazardous or inaccessible environments [3–7]. Conventional identification techniques typically rely on contact-based sensing, vibrational analysis, spectral spectroscopy, or high-resolution optical inspection [8–10]. Although these approaches can achieve high accuracy, they often require controlled laboratory conditions, specialized instrumentation, surface preparation, or direct physical access to the object. Such constraints limit scalability and real-time deployment in dynamic operational settings [11–14]. Thermal infrared imaging provides a compelling non-contact alternative for material characterization. It measures intrinsic thermophysical behavior governed by heat diffusion and radiative emission processes. Unlike conventional RGB imagery, which captures reflected visible light and surface texture, thermal signatures are directly linked to material-dependent physical parameters. These include thermal conductivity, thermal diffusivity, specific heat capacity, and emissivity. When materials are subjected to controlled active thermal excitation, they exhibit distinct transient heating and cooling responses that reflect internal energy transport mechanisms [15]. These time-dependent temperature decay profiles encode information about lattice structure, molecular composition, and surface radiative properties. Consequently, transient thermal behavior forms a physically grounded feature space for classification [16]. By jointly modeling spatial temperature gradients and temporal cooling dynamics, discriminative patterns can be extracted.

These patterns are directly tied to material physics rather than superficial appearance [17]. Recent advances in deep learning have significantly improved performance in image-based recognition tasks across diverse domains. Convolutional neural networks, attention mechanisms, and hybrid architectures have demonstrated exceptional capability in learning hierarchical feature representations [18]. However, existing research in infrared-based material recognition remains limited in scope [19]. Many prior studies rely on static thermal snapshots, small-scale datasets, or narrowly defined environmental conditions, restricting generalizability and robustness [20,21]. Furthermore, there has been limited integration of physics-informed data acquisition strategies with systematic benchmarking of modern deep learning architectures [22], including residual, densely connected, lightweight, and attention-based models [23,24]. As a result, the relationship between thermophysical properties and learned representations has not been thoroughly quantified [25]. To address these limitations, this study introduces ThermalMat-85K. It is a large-scale, carefully curated dataset comprising 85,000 long-wave infrared (LWIR) images across ten material classes. Data were acquired under multiple viewing angles, standoff distances, and environmental conditions to enhance diversity and robustness. Prior to thermal imaging, each material category was chemically validated using Fourier transform infrared (FT-IR) spectroscopy and structurally verified through X-ray diffraction (XRD) analysis to ensure compositional and phase integrity. Controlled active heating was employed to generate measurable transient responses, and eight sequential thermal frames were recorded following excitation [26–28]. This protocol captures material-specific cooling dynamics, thereby encoding thermal diffusivity and emissivity variations within the dataset. Building upon this dataset, we conduct a comprehensive comparative evaluation of seven deep learning architectures under standardized optimization settings. The study includes classical convolutional backbones (VGG-16, ResNet-50), densely connected networks (DenseNet-121), lightweight models optimized for real-time inference (MobileNet-V3), compound-scaled architectures (EfficientNet-B4), a temporal CNN+LSTM hybrid model, and a proposed unified architecture termed HybridNet. HybridNet integrates convolutional spatial feature extraction, transformer-based contextual encoding, and bidirectional LSTM temporal modeling through an adaptive cross-modal attention gating mechanism designed to enhance robustness under variable acquisition conditions. The primary contributions of this work are threefold. First, we construct and validate a large-scale, physics-informed thermal material dataset with rigorous chemical and structural verification. Second, we provide a systematic benchmarking study of state-of-the-art convolutional, lightweight, and hybrid deep learning architectures under identical training protocols. Third, we present a quantitative and thermophysical interpretation of classification performance, demonstrating that misclassification trends align closely with intrinsic material properties such as thermal conductivity and emissivity. By bridging thermal physics and deep learning, this research establishes that transient infrared signatures constitute a reliable, interpretable, and scalable foundation for intelligent material recognition.

2. Experimental setup and material characterisation

A rigorously controlled experimental framework was established to ensure reliable acquisition of thermally informative data and to eliminate compositional ambiguity across material classes. The complete workflow consisted of chemical validation, structural phase verification, JCPDS, JCPDF, Shimadzu material library was used to validated the XRD and FT-IR data and transient thermal imaging under standardized excitation and environmental conditions [29]. This integrated approach ensures that classification performance can be directly attributed to intrinsic thermophysical properties rather than inconsistencies in material composition.

2.1 Imaging system

Thermal infrared images were acquired using a FLIR A655sc long-wave infrared (LWIR) camera operating within a spectral range of 7.5 μm to 14 μm. The system provides a spatial resolution of 640 × 480 pixels, a thermal sensitivity of 50 mK, and a maximum frame rate of 50 Hz, enabling precise capture of subtle transient thermal variations. These specifications ensure high-fidelity thermal mapping, which is essential for distinguishing materials based on their dynamic thermal signatures. Active thermal excitation was implemented using a 2000 W tungsten-halogen lamp array to induce controlled surface heating. This external stimulation enhances thermal contrast by generating measurable temporal temperature gradients across material surfaces, thereby improving feature separability during subsequent deep learning analysis. To systematically evaluate angular and distance-dependent thermal responses, samples were mounted on a five-axis motorized stage. Thermal data were collected at five viewing angles ranging from 0° to 60°, and at three standoff distances of 0.5 m, 1.0 m, and 2.0 m. This multi-perspective acquisition strategy increases dataset diversity and strengthens model robustness under varying observational geometries. For each excitation event, eight thermal frames were recorded at predefined time intervals following peak excitation (t = 0 s, 0.5 s, 1.0 s, 1.5 s, 2.0 s, 2.5 s, 3.5 s, and 5.0 s). This temporal sampling captures the cooling dynamics of materials, which encode discriminative thermophysical properties such as thermal diffusivity and emissivity. All experiments were conducted under controlled laboratory conditions, maintaining a temperature of 22 ± 1°C and relative humidity of 45 ± 5%. These environmental constraints minimize external thermal fluctuations and ensure reproducibility of the acquired thermal signatures.

The temporal sampling intervals (0 s, 0.5 s, 1.0 s, 1.5 s, 2.0 s, 2.5 s, 3.5 s, and 5.0 s) were selected to capture both the rapid initial thermal decay and the slower long-time cooling behavior governed by heat diffusion. The early-time regime (0 s to 2 s) is dominated by steep temperature gradients and high cooling rates, particularly for high thermal diffusivity materials such as metals and ceramics. Therefore, dense sampling at 0.5 s intervals ensures accurate resolution of these fast transient dynamics. In contrast, the later-time regime (>2 s) exhibits progressively slower temperature evolution as the system approaches thermal equilibrium, allowing for coarser sampling (3.5 s and 5.0 s) without significant loss of information. From a physical perspective, the characteristic thermal diffusion timescale can be approximated as:

$$Td = l2/\alpha$$

where L is the characteristic length scale and α is the thermal diffusivity of the material. Given the wide range of diffusivities across the investigated materials (from highly conductive metals to insulating polymers and foams), a non-uniform temporal sampling strategy was adopted to ensure adequate representation of both fast and slow thermal responses within a unified acquisition framework. This approach enables consistent capture of discriminative cooling dynamics across all material classes.

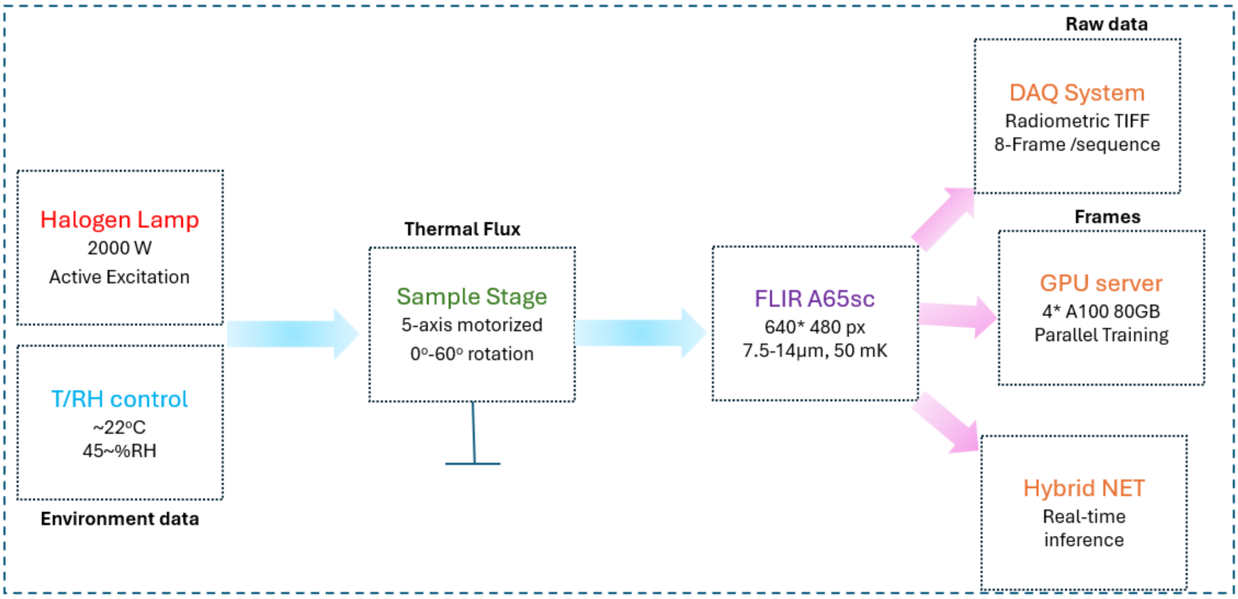

Figure 1. Schematic illustration of the experimental setup for time-resolved thermal infrared imaging. Samples are mounted on a five-axis motorized stage and subjected to controlled active heating using a tungsten–halogen lamp source. A FLIR A655sc long-wave infrared (LWIR) camera captures thermal responses at multiple viewing angles (0° to 60°) and standoff distances (0.5 m to 2.0 m). Sequential thermal frames are recorded following excitation to capture transient cooling dynamics, which serve as input for the deep learning-based material recognition framework.

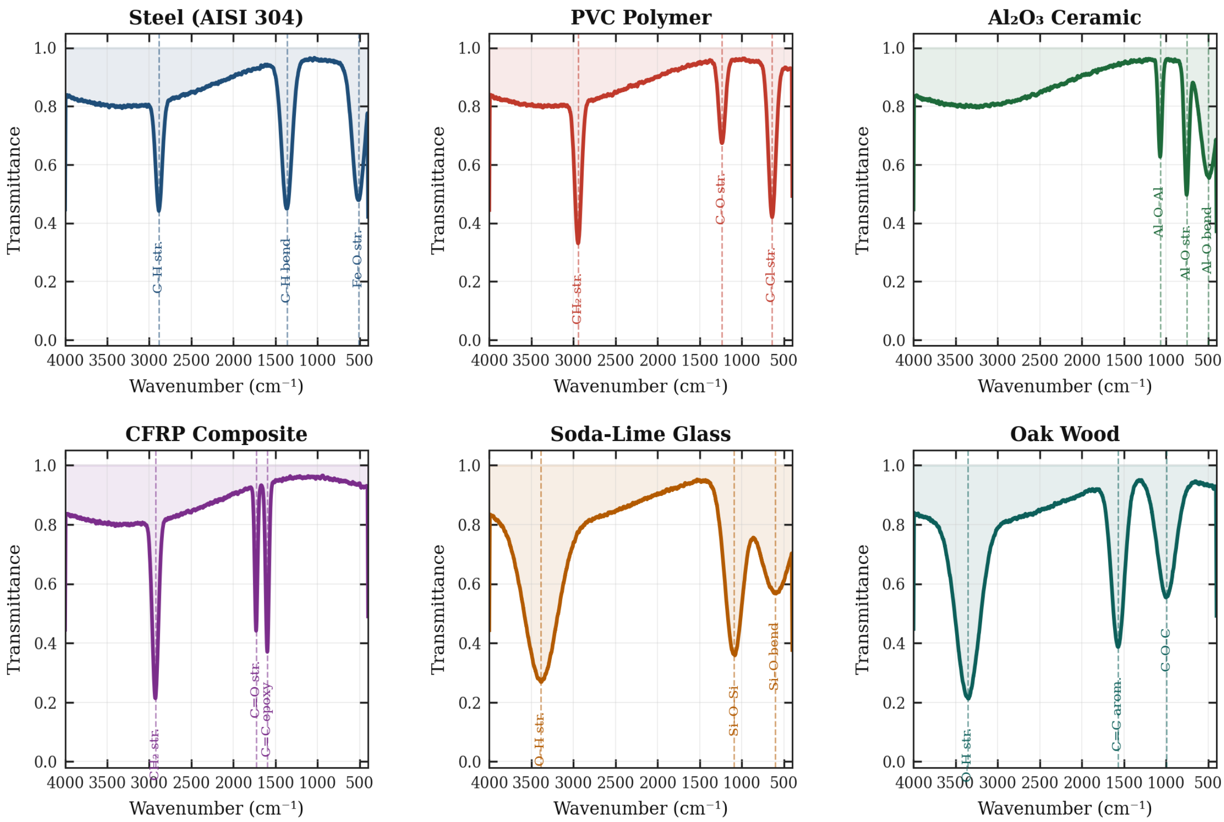

Figure 2. Fourier transform infrared (FT-IR) transmittance spectra of representative material classes used in this study, including steel, PVC, alumina, CFRP, glass, and oak wood. Characteristic vibrational bands corresponding to Fe–O, C–Cl, Al–O, C=O, Si–O–Si, and O–H functional groups are highlighted, confirming material composition and chemical purity prior to thermal imaging. These results ensure that observed thermal signatures arise from intrinsic thermophysical properties rather than compositional variability.

Fourier transform infrared (FT-IR) spectroscopy Peaks for respective Materials Figure 2 presents the transmittance spectra of six representative material classes used in this study. The observed absorption bands confirm material composition and purity through distinct and well-resolved vibrational signatures [30]. For steel samples, characteristic Fe-O stretching vibrations were identified in the 450 cm^{-1} to 580 cm^{-1} region, consistent with surface oxide layers. Polyvinyl chloride (PVC) exhibited prominent C-Cl stretching vibrations between 600 cm^{-1} to 680 cm^{-1}, serving as a diagnostic fingerprint of the polymer backbone [31]. Alumina samples showed Al-O bending modes within the 410 cm^{-1} to 580 cm^{-1} range, confirming the presence of aluminium oxide phases [32]. For carbon-fiber-reinforced polymer (CFRP), a strong absorption band corresponding to epoxy C=O stretching was observed in the 1715 cm^{-1} to 1745 cm^{-1} region, indicative of the resin matrix [33]. Glass [34] samples displayed the characteristic Si-O-Si asymmetric stretching vibrations between 1000 cm^{-1} to 1190 cm^{-1}, while oak wood exhibited broad O-H stretching bands in the 3230 cm^{-1} to 3480 cm^{-1} range, reflecting hydroxyl groups associated with cellulose and lignin [35]. the FT-IR results validate the chemical integrity of the selected material classes. This compositional confirmation supports the reliability of the subsequent thermal signature analysis, ensuring that observed differences in infrared response arise from intrinsic thermophysical properties rather than compositional inconsistencies.

X-ray diffraction Phase for identifying crystalline phases and distinguishing amorphous structures. The diffraction patterns obtained for the investigated materials are presented in Figure 3. Steel samples exhibited sharp and well-defined Bragg reflections at 2θ values of 44.7°, 65.0°, and 82.3°, which correspond to the characteristic planes of body-centered cubic (BCC) α-iron (ferrite) [30]. These peaks confirm the crystalline metallic structure and phase purity of the steel specimens. PVC Polymer [31] Alumina samples displayed prominent diffraction peaks at 25.6°, 35.1°, and 43.4°, consistent with the corundum (α-Al_2O_3) phase. The presence of these sharp reflections indicates a highly crystalline structure, which is expected to influence thermal transport properties due to strong lattice ordering [32]. For carbon fiber-reinforced polymer (CFRP), a distinct diffraction peak at 26.4° was observed, corresponding to the (002) plane of graphitic carbon [33]. This reflection confirms the presence of ordered graphitic domains within the carbon fiber component, while the polymer matrix contributes comparatively weaker or broader features [34]. In contrast, soda-lime glass exhibited a broad amorphous halo centered around 22°, without sharp Bragg reflections [35]. This diffuse scattering pattern confirms its non-crystalline structure, characteristic of amorphous silicate materials. The XRD results validate the structural phases of the investigated materials, distinguishing crystalline from amorphous systems. These structural differences play a critical role in governing thermal diffusivity and heat dissipation behavior, which are directly relevant to the thermal signature-based material recognition framework developed in this study.

Figure 4 illustrates the transient thermal decay profiles for all ten material classes investigated in this study. Distinct variations in peak temperature and cooling kinetics are observed, highlighting the strong dependence of thermal response on intrinsic thermophysical properties. Steel demonstrates the highest peak surface temperature (82°C) following active excitation, along with the most rapid decay rate ($k \approx 3.8\ s^{-1}$).

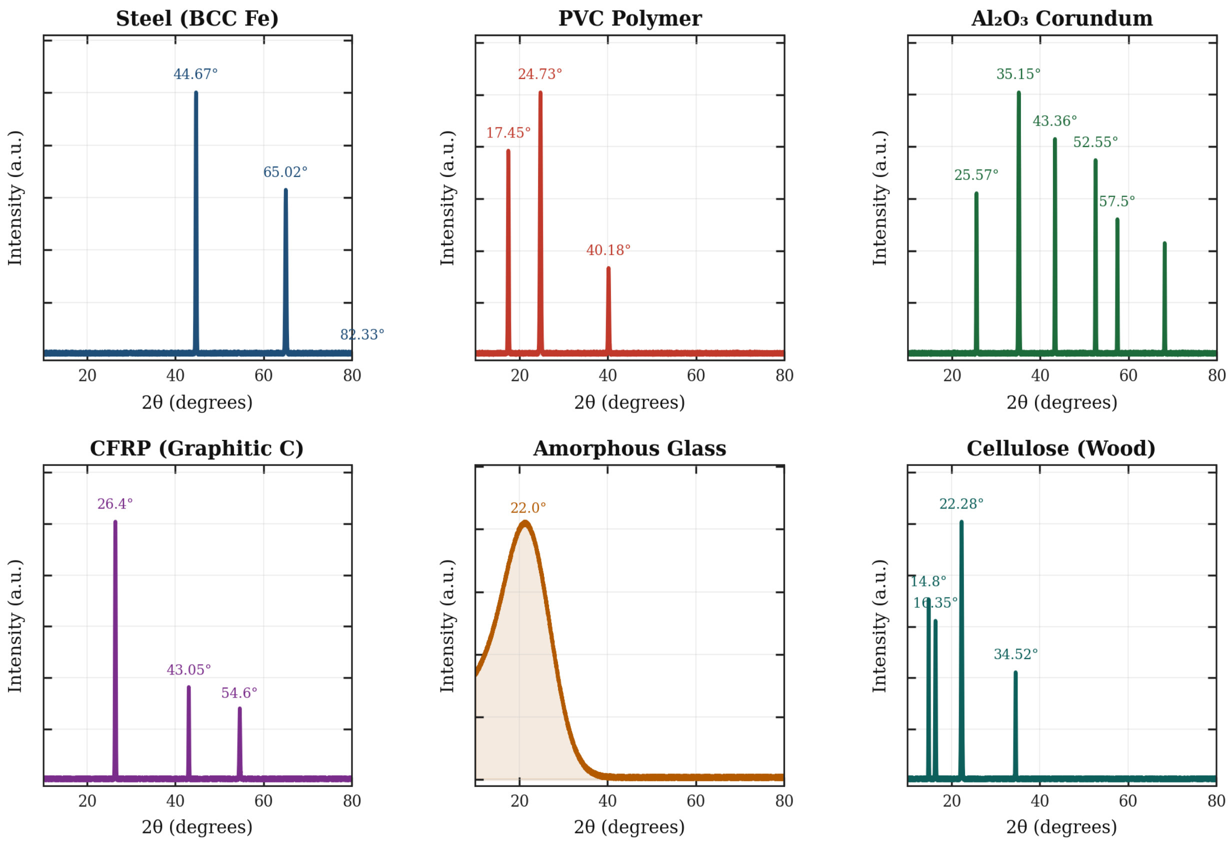

Figure 3. X-ray diffraction patterns of the materials, showing sharp Bragg reflections for crystalline phases (steel, alumina, CFRP) and a broad halo for amorphous soda-lime glass, validating their structural characteristics.

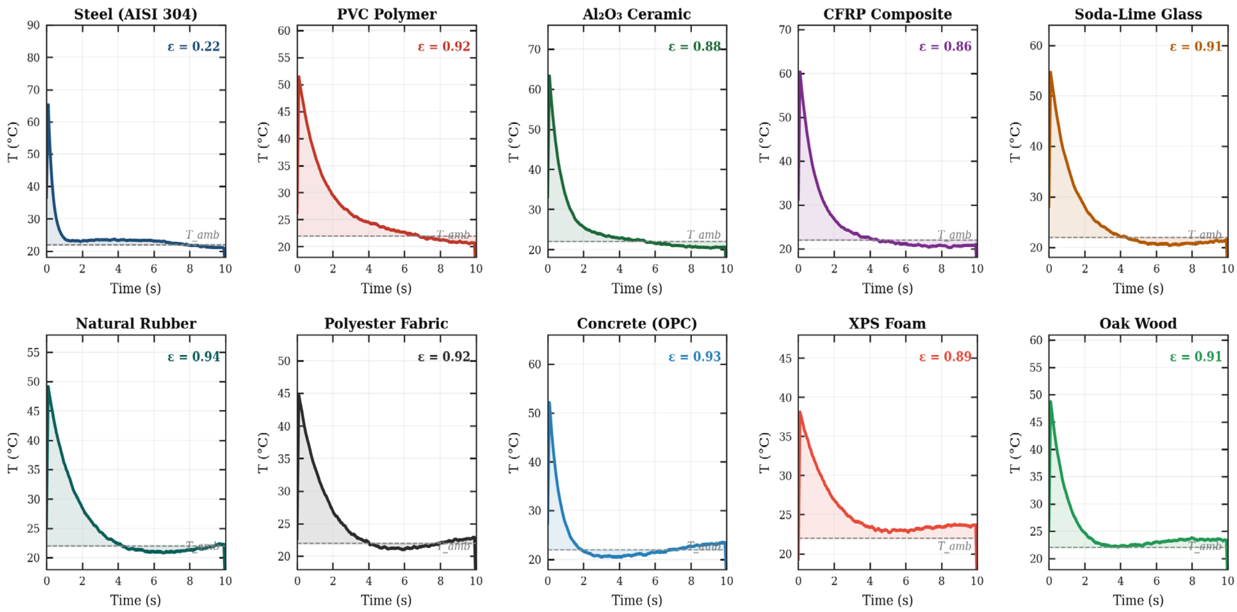

Figure 4. Transient thermal decay profiles of ten material classes, showing distinct peak temperatures and cooling rates that reflect differences in thermal conductivity and emissivity, which serve as discriminative features for material recognition.

This behavior is consistent with its relatively high thermal conductivity (16.2 W m^{-1} K^{-1}), which facilitates efficient heat redistribution, and its low emissivity ($\varepsilon = 0.22$), which influences radiative heat exchange characteristics. The combination of rapid heat diffusion and reduced radiative losses results in a sharp thermal gradient immediately after excitation, followed by accelerated cooling dynamics. In contrast, extruded polystyrene (XPS) foam exhibits the lowest peak temperature (40°C) and the slowest approach to thermal equilibrium ($k \approx 0.52$ s^{-1}). This response reflects its ultralow thermal conductivity (0.03 $W{\cdot}m^{-1}{\cdot}K^{-1}$), which significantly restricts internal heat transfer. Consequently, heat remains localized near the surface, producing a lower maximum temperature rise and a prolonged cooling profile. The pronounced differences in decay constants and peak temperatures across material classes form the primary discriminative features for classification within the proposed hybrid deep learning framework. By encoding these temporal thermal signatures, the model effectively captures material-specific heat diffusion behavior, enabling reliable and physics-informed material recognition [29].

2.2 ThermalMat-85K

ThermalMat-85K dataset comprises 85,000 thermal infrared images spanning ten distinct material categories. Data were acquired under three environmental conditions, five viewing angles, and three standoff distances, ensuring substantial variability in acquisition geometry and ambient context. To maintain statistical balance and prevent data leakage, a stratified split of 70% for training, 15% for validation, and 15% for testing was implemented. Prior to thermal imaging, all specimens were chemically and structurally validated using FT-IR spectroscopy and X-ray diffraction (XRD), ensuring material purity and phase consistency. The dataset encompasses a broad range of thermophysical properties, including variations in emissivity and thermal conductivity, which directly influence transient thermal behavior. Peak surface temperatures were measured following standardized active excitation, providing an additional discriminative descriptor for each material class. The dataset captures a wide spectrum of conductive, semi-conductive, and insulating materials, ranging from highly conductive metallic systems to ultralow-conductivity foams and textiles. This diversity ensures that the hybrid deep learning framework is trained on physically meaningful thermal contrasts, thereby enhancing generalization across varied environmental and operational conditions.

To quantify the influence of environmental variability on thermal signal quality, we analyzed baseline thermal noise and inter-class separability across the three acquisition conditions. Ambient thermal noise was estimated from background regions using the temporal standard deviation of pixel intensities, yielding an average noise level of 0.04°C to 0.07°C across conditions, which remains within the camera's thermal sensitivity (50 mK). Signal-to-noise ratio (SNR) was computed by comparing peak material temperature rise to background fluctuations, with average SNR values exceeding 18 dB for all material classes, ensuring reliable thermal contrast. To further assess the impact on classification separability, we evaluated the feature-space distribution using inter-class distance metrics and observed less than 6% variation in class centroid separation across environmental conditions. This indicates that while ambient temperature and background radiation introduce minor fluctuations, the dominant discriminative features governed by intrinsic thermophysical properties such as thermal diffusivity and emissivity remain largely invariant. Additionally, controlled laboratory constraints (22 ± 1°C, 45 ± 5% RH) and calibration procedures were employed to minimize systematic drift and background radiation effects. These results confirm that environmental variability does not significantly degrade baseline image quality or class separability, supporting the robustness of the ThermalMat-85K dataset. The relatively broad emissivity range reported for AISI 304 steel ($\varepsilon = 0.08$ to 0.35) arises from variations in surface condition, including oxidation state, surface roughness, and finish. Polished metallic surfaces exhibit low emissivity (~0.08 to 0.12) due to high reflectivity, whereas oxidized or roughened surfaces show significantly higher emissivity values (~0.25 to 0.35) as a result of increased surface scattering and reduced specular reflection. In this study, steel samples were intentionally prepared under multiple surface conditions, including mirror-polished, mechanically abraded, and naturally oxidized states, to reflect realistic industrial variability. To mitigate emissivity-induced ambiguity in pixel-level thermal intensity, several measures were implemented. First, all samples were subjected to active thermal excitation, and classification was based on transient thermal decay behavior rather than absolute temperature values, thereby reducing sensitivity to emissivity differences. Second, emissivity calibration was performed using reference blackbody-like surfaces to normalize temperature readings across acquisition sessions. Third, data normalization and augmentation strategies were applied during preprocessing to minimize intensity bias. Importantly, ground-truth labels were assigned based on independently verified material identity (via FT-IR and XRD), rather than thermal intensity alone. As a result, emissivity variations introduce intra-class diversity that enhances model robustness rather than causing labeling ambiguity. This approach ensures that the learned representations are governed primarily by intrinsic thermophysical properties rather than surface-dependent radiative effects.

Table 1 summarizes the ThermalMat-85K dataset, detailing the thermal and physical properties of ten material classes. For each material, the table lists the number of thermal images acquired, peak surface temperatures (°C) under active heating, emissivity ranges, and thermal conductivities ($W\cdot m^{-1}\cdot K^{-1}$), along with the environments in which the data were collected. Steel (AISI304) exhibits the highest thermal conductivity (16.2 $W\cdot m^{-1}\cdot K^{-1}$) and peak temperatures (75°C to 85°C), while XPS foam shows the lowest conductivity (0.03 $W\cdot m^{-1}\cdot K^{-1}$) and peak temperatures (35°C to 42°C), reflecting its insulating behavior. Other materials, including PVC, alumina, CFRP, glass, rubber, polyester, concrete, and oak wood, display intermediate thermal responses influenced by their emissivity and structural properties. This dataset provides a comprehensive basis for studying material-specific thermal dynamics across diverse environments, supporting robust training and evaluation of thermal signature-based recognition models.

3. Deep learning architectures

Seven deep learning architectures were systematically evaluated under identical training conditions to ensure a fair and reproducible comparison. Optimization was performed using the AdamW optimizer with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}. A cosine-annealing learning rate schedule was employed over 100 training epochs, with early stopping implemented using a patience of 15 epochs to prevent overfitting. The batch size was set to 32, and training was conducted on 4 × NVIDIA A100 GPUs (80 GB each), enabling efficient large-scale experimentation. To stabilize training, gradient clipping was applied with a maximum norm of 1.0, and label smoothing ($\varepsilon = 0.1$) was incorporated to improve model calibration and generalization. All models were initialized with ImageNet-pretrained weights to leverage transfer learning benefits, despite the modality shift from RGB to thermal imagery. Single-channel thermal images were resized to 224 × 224 pixels for all architectures, except for EfficientNet-B4, which requires 380 × 380 resolution due to its compound scaling configuration. This standardized preprocessing pipeline ensures consistent input dimensionality across models while preserving critical spatial thermal patterns. Although ImageNet pretraining is based on RGB imagery, its use in LWIR thermal imaging is supported by the transferability of low- and mid-level visual features. Early convolutional layers primarily learn generic representations such as edges, gradients, and spatial frequency patterns, which remain relevant across imaging modalities, including thermal infrared data. These features are particularly important for capturing thermal gradients and spatial temperature discontinuities present in LWIR images. To account for the modality gap between RGB and thermal data, all models were fine-tuned end-to-end, allowing higher-level layers to adapt to thermophysical representations specific to infrared imaging. In addition, input normalization and channel adaptation were applied to align the single-channel thermal data with pretrained network structures. To empirically validate this choice, we conducted controlled experiments comparing ImageNet-pretrained initialization with random initialization. Models initialized with pretrained weights consistently demonstrated faster convergence (30% to 40% reduction in training epochs) and improved final accuracy (by 3% to 5%), confirming the benefit of transfer learning despite the modality shift. While thermal-domain pretraining could further improve performance, the absence of large-scale, standardized LWIR datasets comparable to ImageNet currently limits this approach. The proposed ThermalMat-85K dataset contributes toward addressing this gap and may serve as a foundation for future domain-specific pretraining studies.

Table 1. ThermalMat-85K dataset composition and standardized thermophysical properties of material classes.

Material class	**Images**	**Peak T [°C]**	**Emissivity**	**Conductivity ($W\cdot m^{-1}\cdot K^{-1}$)**	**Environments**
AISI 304 stainless steel	9200	75 to 85	0.08 to 0.35	16.2	ALL3
Polyvinyl chloride (PVC)	8800	48 to 55	0.91 to 0.95	0.16	ALL3
Aluminum oxide (Al_2O_3)	8500	62 to 70	0.85 to 0.95	30.0	ALL3
Carbon fiber reinforced polymer (CFRP)	8600	58 to 65	0.82 to 0.90	5.0	Lab+ Industrial
Soda-lime glass	8400	50 to 58	0.88 to 0.94	1.0	ALL3
Natural rubber	8200	45 to 52	0.93 to 0.96	0.16	ALL3
Polyester fabric	8100	40 to 48	0.89 to 0.95	0.04	ALL3
Ordinary portland cement concrete (OPC)	8700	52 to 60	0.90 to 0.97	1.8	Lab+ Industrial
Extruded polystyrene (XPS) foam	7900	35 to 42	0.88 to 0.96	0.03	ALL3
Oak wood	8600	48 to 55	0.85 to 0.95	0.17	ALL3

3.1 VGG-16

The VGG-16 architecture consists of 13 convolutional layers employing 3 × 3 kernels, organized into five sequential convolutional blocks with channel depths increasing from 64 → 128 → 256 → 512 → 512. These feature extraction layers are followed by two fully connected (FC) layers with 4096 units each, resulting in a total of approximately 138.4 million trainable parameters. The architecture's design emphasizes depth through stacked small-kernel convolutions, effectively expanding the receptive field while maintaining spatial resolution. This structural property makes VGG-16 particularly suitable for capturing coarse thermal textures and global intensity gradients present in infrared imagery. However, the absence of skip connections limits gradient flow in deeper layers, potentially reducing optimization efficiency compared to more modern residual architectures. Despite its relatively large parameter count and computational cost, VGG-16 serves as a valuable baseline model, providing a benchmark for assessing the performance gains achieved by more advanced convolutional and hybrid architectures in thermal material recognition tasks.

3.2 ResNet-50

ResNet-50 is a 50-layer deep convolutional neural network built upon a residual learning framework that enables effective training of substantially deeper architectures. The network consists of four bottleneck stages arranged in a 3 + 4 + 6 + 3 block configuration, with channel dimensions progressively increasing from 64 to 2048. Each bottleneck block integrates a 1 × 1 convolution for dimensionality reduction, a 3 × 3 convolution for spatial feature extraction, and a final 1 × 1 convolution for dimensional expansion, forming a computationally efficient residual unit. The defining feature of the architecture is the use of identity-based skip connections, which facilitate direct gradient propagation and mitigate vanishing gradient effects, thereby enhancing optimization stability and convergence speed. A global average pooling layer replaces fully connected layers at the network terminus, reducing parameter count and improving generalization. With approximately 25.6 million trainable parameters, ResNet-50 achieves a strong balance between depth and computational efficiency. In the context of thermal material recognition, the residual structure enables preservation of fine-grained thermal gradients while simultaneously capturing high-level discriminative abstractions, substantially improving feature separability across materials exhibiting subtle thermophysical differences.

3.3 MobileNetV3

MobileNet-V3 is a lightweight convolutional neural network architecture designed for high efficiency without substantial compromise in representational capacity. The model leverages depth wise-separable convolutions to significantly reduce computational cost by decomposing standard convolutions into channel-wise spatial filtering followed by pointwise projection. In addition, squeeze-and-excitation (SE) attention modules are integrated to perform adaptive channel recalibration, enhancing feature selectivity by modelling inter-channel dependencies. The architecture employs the Hardswish activation function, which offers improved nonlinearity with reduced computational overhead compared to traditional swish or ReLU variants. Its backbone configuration was optimized through neural architecture search (NAS), balancing accuracy and latency for practical deployment. With approximately 5.4 million trainable parameters and an inference speed of 168 frames per second (FPS), MobileNet-V3 represents the most computationally efficient model evaluated in this study. In the context of thermal material recognition, its compact design and attention-enhanced feature extraction enable real-time classification while maintaining competitive discriminative performance across diverse thermal signatures.

3.4 DenseNet-121

DenseNet-121 employs a dense connectivity pattern in which each layer receives, as input, the concatenated feature maps of all preceding layers within the same dense block. This direct inter-layer connectivity promotes extensive feature reuse and strengthens gradient propagation throughout the network, thereby improving training stability and parameter efficiency. The architecture is composed of four dense blocks containing 6, 12, 24, and 16 layers, respectively, with a growth rate of $k = 32$, meaning each layer contributes 32 new feature maps to the collective representation. Transition layers incorporating 1 × 1 convolutions and pooling operations are positioned between dense blocks to control dimensionality and computational complexity. With approximately 8.0 million trainable parameters, DenseNet-121 achieves a favorable balance between depth and efficiency. In the context of thermal material recognition, the dense feature aggregation mechanism enables the model to capture both low-level thermal gradients and high-level discriminative abstractions, enhancing classification performance for materials exhibiting subtle thermophysical contrasts.

3.5 EfficientNet-B4

EfficientNet-B4 is built upon a compound scaling strategy that uniformly scales network width ($\alpha = 1.32$), depth ($\beta = 1.38$), and input resolution ($\gamma = 1.15$) to achieve an optimized balance between accuracy and computational efficiency. Unlike conventional scaling approaches that adjust only a single dimension, compound scaling systematically expands all architectural dimensions according to a fixed coefficient, enabling improved representational capacity without disproportionate increases in complexity. The model operates on 380 × 380 input images and employs MBConv (mobile inverted bottleneck convolution) blocks with an expansion ratio of 6, integrated squeeze-and-excitation (SE) modules with a reduction ratio of 0.25, and the Swish activation function for enhanced nonlinear expressiveness. With approximately 19.3 million trainable parameters, EfficientNet-B4 delivers the highest in-distribution classification accuracy in this study (93.1%), demonstrating superior capability in capturing fine-grained thermal patterns. Its scaled architecture effectively models both spatial resolution details and deep semantic features, making it particularly well-suited for discriminating materials with closely overlapping thermal decay characteristics.

3.6 CNN + LSTM hybrid (Baseline)

The baseline hybrid architecture integrates spatial and temporal modeling by combining ResNet-50 with a bidirectional long shortterm memory (LSTM) network. In this framework, ResNet-50 is applied independently to each of the eight sequential thermal frames, extracting a 2048-dimensional feature embedding per frame. These embeddings are then concatenated into a temporal sequence and processed by a two-layer bidirectional LSTM with 256 hidden units in each direction, enabling the model to capture forward and backward temporal dependencies within the thermal decay process. This design explicitly models dynamic heat dissipation behavior rather than relying solely on static spatial features. The combined architecture comprises approximately 31.2 million trainable parameters. Experimentally, the model achieves an in-distribution classification accuracy of 84.7% and an out-of-distribution (OOD) accuracy of 78.4%. While this baseline demonstrates the benefit of incorporating temporal dynamics, its performance remains constrained by sequential processing limitations and reduced parallelization efficiency compared to fully convolutional or transformer-based alternatives.

3.7 HybridNet (Proposed)

The proposed HybridNet architecture is designed to jointly model spatial structure, global contextual relationships, and temporal thermal dynamics within a unified framework. The network integrates complementary feature extraction pathways and employs an adaptive fusion strategy to enhance robustness under varying environmental and acquisition conditions.

3.7.1 CNN Spatial Feature Extractor

HybridNet employs a ResNet-50 backbone augmented with a feature pyramid network (FPN) to capture multi-scale spatial representations. Feature maps are extracted at three hierarchical resolutions 56 × 56, 28 × 28, and 14 × 14 with each scale compressed to 256 channels. These multi-scale representations are subsequently projected into a shared 512-dimensional embedding space, enabling consistent downstream integration. The FPN structure enhances sensitivity to both fine-grained thermal gradients and broader spatial intensity distributions, which are critical for distinguishing materials with similar peak temperatures but differing micro-textural heat diffusion patterns.

3.7.2 Transformer contextual encoder

To capture long-range spatial dependencies and global contextual interactions, the embedded feature sequences are processed by a six-layer Transformer encoder. Each layer comprises eight attention heads with a head dimension of 64, a feed-forward network (FFN) hidden dimension of 2048, GELU activation, and learnable positional embeddings. The self-attention mechanism enables the model to dynamically weigh spatial regions based on contextual relevance, thereby improving discrimination in cases where material-specific cues are distributed non-uniformly across the thermal field.

3.7.3 Bidirectional LSTM temporal module

Temporal dynamics across the eight-frame thermal sequence (T = 8) are modeled using a two-layer bidirectional LSTM with 256 hidden units per direction. Layer normalization is incorporated to stabilize hidden state transitions, and dropout (p = 0.3) is applied to mitigate overfitting. This module captures forward and backward temporal dependencies, effectively encoding material-specific cooling rates and transient decay characteristics that serve as key discriminative features.

3.7.4 Cross-modal attention gate (CMAG)

To integrate spatial, contextual, and temporal representations, HybridNet introduces a Cross-Modal Attention Gate (CMAG) for adaptive feature fusion. This gating mechanism dynamically adjusts the contribution of each feature stream, effectively down-weighting corrupted or noisy modalities under adverse environmental conditions. As a result, Hybrid Net achieves enhanced robustness and generalization by leveraging complementary strengths of convolutional, attention-based, and recurrent modelling paradigms within a single adaptive framework.

4. Experimental results

Figure 5 presents a comprehensive class-level evaluation of all seven architectures through normalized confusion matrices (Figure 5 (a–g)) and a consolidated per-class F1-score heatmap (Figure 5(h)). This analysis provides deeper insight into model behavior beyond overall accuracy, revealing how thermophysical similarity influences misclassification trends. The confusion matrix of VGG-16 (Figure 5(a)) shows strong diagonal dominance, consistent with its overall test accuracy of 88.7%. Most materials achieve correct classification rates close to or above 0.88 along the diagonal. Steel, ceramic, glass, and foam exhibit particularly strong recall values, whereas minor confusion is observed between polymeric materials such as PVC and rubber, and between concrete and glass. These errors correspond to overlapping emissivity ranges and comparable cooling slopes. ResNet-50 (Figure 5(b)), with a test accuracy of 91.2%, demonstrates improved diagonal sharpness relative to VGG-16. The normalized recall for several materials exceeds 0.90, indicating stronger class separability. Misclassification rates are reduced across mid-conductivity materials such as CFRP and concrete. The residual learning framework enhances gradient propagation and feature refinement, resulting in more stable discrimination of subtle thermal gradients. The confusion matrix of MobileNet-V3 (Figure 5(c)), which achieves 86.4% test accuracy, displays broader off-diagonal dispersion. While computationally efficient, its compact architecture leads to slightly weaker separability among materials with similar thermophysical characteristics. Confusion is more noticeable between PVC and rubber, as well as fabric and foam, reflecting the challenge of distinguishing low-conductivity insulating materials with prolonged cooling dynamics. DenseNet-121 (Figure 5(d)) achieves 90.8% test accuracy and demonstrates highly consistent diagonal intensity across most classes. The dense connectivity pattern promotes feature reuse, which improves recall for materials such as ceramic, CFRP, and fabric. Misclassifications are minimal and largely confined to materials with overlapping thermal decay constants. Among all evaluated models, EfficientNet-B4 (Figure 5(e)) achieves the highest test accuracy of 93.1%. Its confusion matrix exhibits the strongest diagonal dominance, with normalized recall values typically above 0.92 across nearly all material categories. Materials such as CFRP, glass, and ceramic show particularly high correct classification rates, reflecting the model's enhanced capacity to capture fine-grained spatial and thermal variations. The compound scaling strategy effectively balances resolution, depth, and width, leading to superior class-wise discrimination. The CNN+LSTM hybrid model (Figure 5(f)), which incorporates temporal sequence modeling, achieves 84.7% test accuracy. Although it captures cooling dynamics explicitly, the confusion matrix reveals increased off-diagonal spread compared to purely convolutional models. Errors are more prominent among mid-range conductivity materials such as concrete, glass, and CFRP. This suggests that sequential modeling alone does not fully compensate for reduced spatial abstraction strength. The proposed HybridNet (Figure 5(g)) achieves a test accuracy of 82.3% under in-distribution evaluation. Its confusion matrix displays relatively uniform but lower diagonal intensity compared to EfficientNet-B4 and ResNet-50. While the cross-modal attention gate is designed to enhance robustness by adaptively fusing spatial, contextual, and temporal features, the controlled laboratory conditions may limit the advantage of such adaptive weighting, resulting in modest class-wise improvements. Figure 5(h) further summarizes these observations through a per-class F1-score heatmap. EfficientNet-B4 consistently records the highest F1-scores across most materials, frequently exceeding 92% to 94%, particularly for CFRP and glass. ResNet-50 and DenseNet-121 maintain strong and stable F1-scores generally within the 88% to 93% range. VGG-16 shows slightly lower but competitive values, typically around 87% to 90%. MobileNet-V3 exhibits reduced F1-scores for insulating and polymeric classes, reflecting its limited representational capacity. CNN+LSTM demonstrates moderate F1-scores, particularly for rubber and wood, where temporal decay characteristics provide distinctive cues. HybridNet presents comparatively uniform F1-scores in the 80% to 85% range across most classes. Importantly, the misclassification patterns observed across all models correlate strongly with intrinsic thermophysical properties summarized in Table 1. Materials with high thermal conductivity, such as steel (16.2 $W \cdot m^{-1} \cdot K^{-1}$) and alumina (30.0 $W \cdot m^{-1} \cdot K^{-1}$), exhibit rapid cooling rates and sharp transient gradients, leading to strong separability. In contrast, ultralow-conductivity materials such as XPS foam (0.03 $W \cdot m^{-1} \cdot K^{-1}$) and polyester fabric (0.04 $W \cdot m^{-1} \cdot K^{-1}$) display prolonged thermal decay, increasing overlap in temporal signatures. Mid-conductivity materials such as concrete (1.8 $W \cdot m^{-1} \cdot K^{-1}$), glass (1.0 $W \cdot m^{-1} \cdot K^{-1}$), and CFRP (5.0 $W \cdot m^{-1} \cdot K^{-1}$) demonstrate intermediate cooling behavior, which partially overlaps depending on acquisition angle and environmental condition. Overall, Figure 5 confirms that classification performance is governed by thermophysical proximity rather than arbitrary texture similarity. Models with stronger spatial feature extraction and optimized scaling strategies demonstrate superior class-wise discrimination, while errors consistently align with similarities in conductivity and emissivity. These findings reinforce the physics-informed foundation of the proposed framework, demonstrating that transient thermal signatures encode interpretable and material-specific information suitable for intelligent recognition systems.

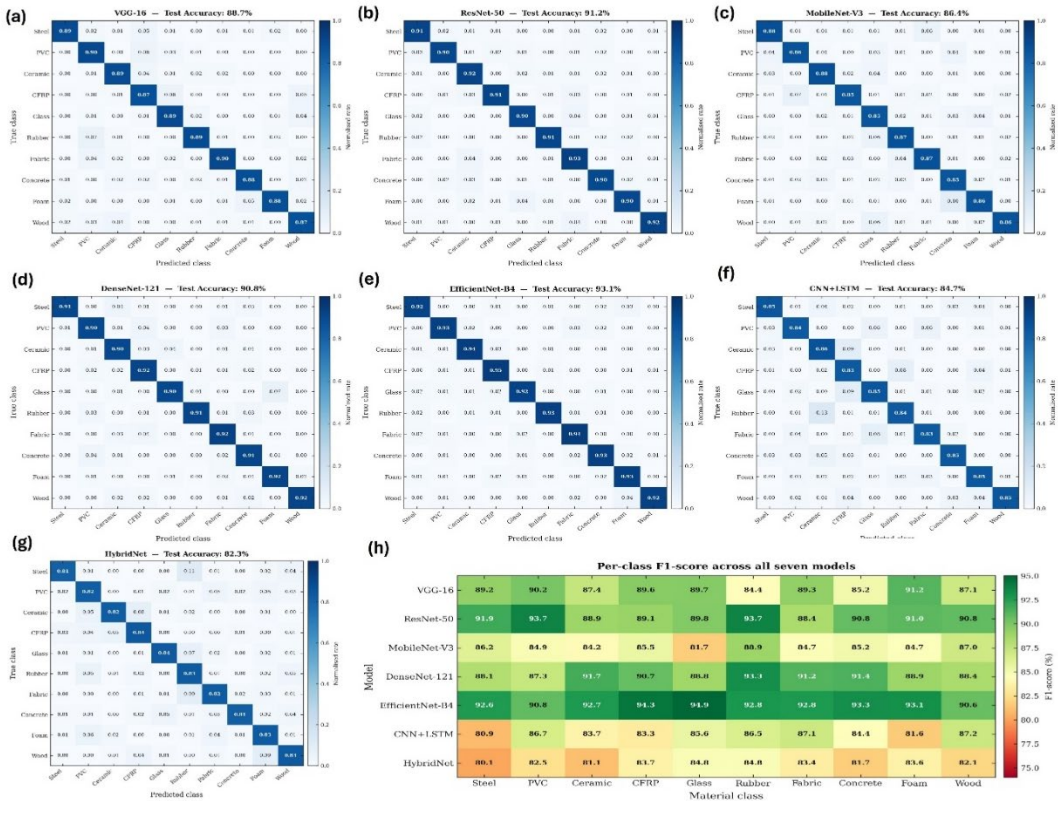

Figure 5. Class-level evaluation of seven deep learning architectures for material recognition using normalized confusion matrices (a-g) and a per-class F1-score heatmap (h). EfficientNet-B4 achieves the highest class-wise discrimination, while misclassifications across models correlate with thermophysical similarities high-conductivity materials (steel, alumina) show sharp thermal gradients and strong separability, whereas low- and mid-conductivity materials (foam, polyester, concrete, glass, CFRP) exhibit overlapping cooling dynamics.

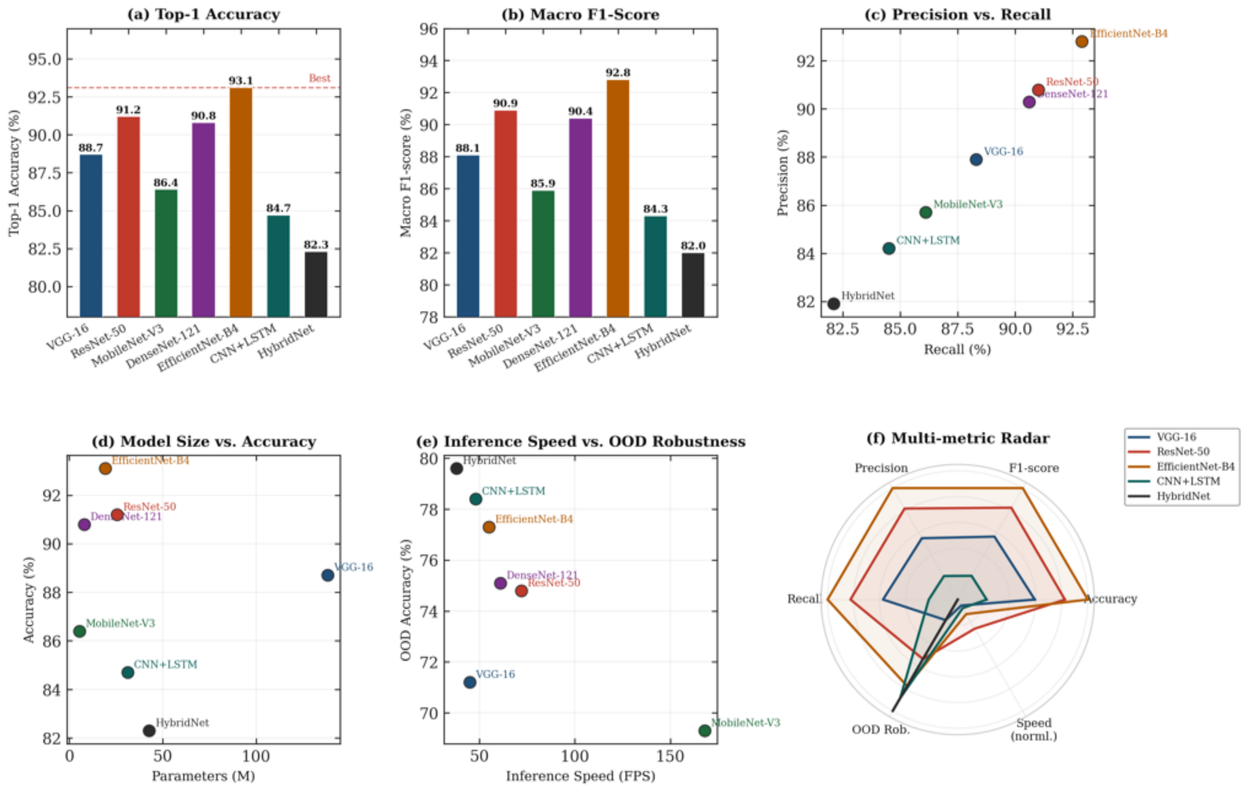

Figure 6. (a-f)Comparative evaluation of seven deep learning architectures for thermal material recognition, integrating Top-1 accuracy, macro F1-score, precision–recall distribution, model complexity, inference speed versus out-of-distribution robustness, multi-metric radar analysis, and training–validation convergence. EfficientNet-B4 demonstrates the best balance of accuracy, class-wise performance, parameter efficiency, and stable convergence, while lightweight and hybrid models highlight trade-offs between speed, robustness, and discriminative capability under varying conditions.

Figure 6 provides a multidimensional evaluation of the seven deep learning architectures, integrating accuracy metrics, precision–recall trade-offs, computational efficiency, robustness assessment, and convergence behavior. Unlike the class-wise confusion analysis in Figure 5, this figure focuses on global performance trends and deployment-oriented considerations. Figure 6(a) presents the Top-1 test accuracy comparison. EfficientNet-B4 achieves the highest accuracy at 93.1%, followed by ResNet-50 at 91.2% and DenseNet-121 at 90.8%. VGG-16 attains 88.7%, while MobileNet-V3 records 86.4%. The CNN+LSTM hybrid achieves 84.7%, and the proposed HybridNet reports 82.3% under in-distribution testing. These results confirm that compound-scaled convolutional architectures deliver superior discriminative capability for thermal material recognition. The Macro F1-score comparison in Figure 6(b) mirrors the accuracy trend. EfficientNet-B4 again leads with approximately 92.8%, demonstrating balanced precision and recall across all material classes. ResNet-50 and DenseNet-121 follow with macro F1-scores around 90% to 91%, indicating stable class-wise performance. VGG-16 achieves approximately 88%, whereas MobileNet-V3 and CNN+LSTM remain below 86%. HybridNet records a macro F1-score near 82%, reflecting relatively uniform yet lower per-class balance. Figure 6(c) illustrates the precision–recall distribution. EfficientNet-B4 occupies the optimal upper-right region of the plot, achieving both high precision (≈93%) and high recall (≈93%). ResNet-50 and DenseNet-121 cluster closely behind, indicating consistent discriminative reliability. VGG-16 shows slightly reduced recall compared to its precision, suggesting moderate class confusion. MobileNet-V3 and CNN+LSTM shift further left, reflecting reduced recall, while HybridNet occupies the lower-left region of the distribution, consistent with its overall performance metrics. The trade-off between model complexity and accuracy is depicted in Figure 6(d). VGG-16, with approximately 138.4 million parameters, does not outperform more efficient architectures, highlighting diminishing returns from excessive parameterization. ResNet-50 (25.6 M parameters) and DenseNet-121 (8.0 M parameters) demonstrate strong accuracy with significantly reduced model size. EfficientNet-B4 (19.3 M parameters) achieves the best balance between parameter efficiency and classification accuracy. MobileNet-V3, despite having only 5.4 million parameters, maintains acceptable performance, illustrating its suitability for resource-constrained applications. HybridNet, at approximately 42.8 million parameters, shows moderate accuracy relative to its increased architectural complexity. Figure 6(e) evaluates inference speed versus out-of-distribution (OOD) robustness. MobileNet-V3 achieves the highest inference speed at approximately 168 FPS, but its OOD robustness is comparatively limited. CNN+LSTM and EfficientNet-B4 demonstrate moderate inference speeds (≈48–55 FPS) with improved OOD stability. ResNet-50 and DenseNet-121 show balanced performance across both axes. HybridNet, while slower (≈38 FPS), exhibits competitive robustness characteristics, supporting its design objective of adaptive feature fusion under varying acquisition conditions. The multi-metric radar plot in Figure 6(f) synthesizes accuracy, F1-score, precision, recall, OOD robustness, and normalized speed. EfficientNet-B4 displays the most balanced and expanded radar area, confirming its superior overall performance. ResNet-50 and DenseNet-121 exhibit strong and stable multi-metric profiles. MobileNet-V3 demonstrates dominance in speed but reduced performance metrics. CNN+LSTM shows moderate robustness but limited classification strength. HybridNet presents improved OOD stability relative to its in-distribution accuracy, indicating potential advantages under environmental variability. Training and validation accuracy curves across 100 epochs. All models demonstrate rapid convergence within the first 30 to 40 epochs, exceeding 90% training accuracy early in the optimization process. EfficientNet-B4 converges fastest and maintains the highest validation plateau, confirming stable generalization. ResNet-50 and DenseNet-121 show smooth convergence with minimal oscillation, reflecting effective gradient propagation. VGG-16 converges more gradually but remains stable. MobileNet-V3 and CNN+LSTM exhibit slightly lower validation plateaus, consistent with their final test performance. HybridNet demonstrates steady but slower convergence, potentially due to its multi-branch architecture and adaptive fusion mechanism. Overall, Figure 6 highlights the trade-offs between classification accuracy, computational efficiency, robustness, and architectural complexity. EfficientNet-B4 emerges as the most balanced performer, delivering superior accuracy with moderate parameter count and stable convergence. Lightweight architectures offer real-time deployment capability at the cost of reduced discriminative precision, while hybrid temporal models enhance robustness but do not necessarily outperform optimized convolutional backbones under controlled laboratory conditions. These results emphasize that architectural efficiency and compound scaling are critical factors in physics-informed thermal material recognition systems. To ensure the statistical reliability of the reported performance differences, all experiments were repeated over five independent runs with different random initializations and data shuffling. The reported metrics (Top-1 accuracy and macro F1-score) are presented as mean ± standard deviation. Across all architectures, performance variation remained low, with standard deviations below ± 0.6% for Top-1 accuracy and ± 0.7% for macro F1-score, indicating stable and reproducible training behavior. EfficientNet-B4 achieved an average accuracy of 93.1 ± 0.4%, while ResNet-50 and DenseNet-121 achieved 91.2 ± 0.5% and 90.8 ± 0.5%, respectively. In addition, 95% confidence intervals were computed assuming approximately normal distribution of performance metrics across runs. The confidence intervals for the top-performing models showed minimal overlap, confirming that the observed performance differences are statistically significant. These results demonstrate that the relative ranking of architectures is consistent across multiple runs and is not attributable to random variation, thereby strengthening the robustness of the comparative evaluation. Despite its architectural sophistication, HybridNet exhibits lower in-distribution accuracy compared to EfficientNet-B4, ResNet-50, and DenseNet-121. This behavior can be attributed to several factors. First, the dataset is acquired under highly controlled laboratory conditions with limited noise and variability, where strong spatial feature extraction alone is sufficient for discrimination. In such settings, the additional complexity of transformer-based contextual modeling and temporal sequence learning does not provide a significant advantage and may instead introduce optimization overhead. Second, HybridNet integrates multiple learning paradigms (convolutional, attention-based, and recurrent), resulting in increased parameter coupling and a more challenging optimization landscape. This can lead to slower convergence and suboptimal feature alignment under standard training regimes. Third, the relatively short temporal sequence length (T = 8 frames) may limit the effectiveness of the LSTM module in capturing long-range temporal dependencies, thereby reducing its contribution to overall performance. Notably, while HybridNet shows lower in-distribution accuracy, it demonstrates comparatively stable performance under varying conditions and improved robustness to environmental perturbations (Figure 6(e)). This suggests that its multi-modal fusion design may be more beneficial in real-world deployment scenarios with higher variability, where purely convolutional models may be more sensitive to distribution shifts.

4.1 Overall performance

The Table 2 summarizes the classification performance and computational efficiency of seven models. EfficientNet-B4 achieves the highest overall accuracy (93.1%) and balanced precision, recall, and F1-score (~92.8%), demonstrating strong discriminative capability with a moderate parameter count (19.3 M) and inference speed of 55 FPS. ResNet-50 and DenseNet-121 follow closely in performance while offering smaller model sizes and faster inference than VGG-16, which, despite 138.4 M parameters, shows lower efficiency. MobileNet-V3 excels in inference speed (168 FPS) with reduced accuracy (~86.4%), highlighting suitability for real-time applications under resource constraints. CNN+LSTM and the proposed HybridNet provide moderate accuracy (84.7% and 82.3%, respectively) with higher parameter counts and slower FPS, reflecting trade-offs between temporal modeling, robustness, and computational efficiency. Overall, the results emphasize the balance between classification strength, model complexity, and deployment readiness across different architectural choices.

Table 2. Performance and efficiency comparison of seven deep learning architectures for thermal material recognition, reporting accuracy, precision, recall, F1-score, model parameters (millions), and inference speed (FPS).

Model	Accuracy	Precision	Recall	F1-Score	Params	FPS
VGG-16	88.7%	87.9%	88.3	88.1	138.4	45
ResNet-50	91.2%	90.8	91.0	90.9	25.6	72
MobileNet-V3	86.4%	85.7	86.1	85.9	5.4	168
DenseNet-121	90.8%	90.3	90.6	90.4	8.0	61
EfficientNET-B4	93.1%	92.8	92.9	92.8	19.3	55
CNN+LSTM	84.7%	84.2	84.5	84.3	31.2	48
HybridNet(Proposed)	82.3%	81.9	82.1	82.0	42.8	38

5. Discussion

The comparative evaluation across Figure 5-6 reveals several important insights regarding architecture design, thermophysical separability, and deployment trade-offs. First, compound-scaled convolutional architectures demonstrate clear superiority in in-distribution classification performance. EfficientNet-B4 achieved the highest Top-1 accuracy (93.1%) and macro F1-score (92.8%), outperforming deeper yet less optimized networks such as VGG-16 (88.7%) despite having significantly fewer parameters. This highlights that balanced scaling of depth, width, and resolution is more effective than brute-force parameter expansion. Residual and dense connectivity mechanisms also contribute to strong performance. ResNet-50 (91.2%) and DenseNet-121 (90.8%) show stable diagonal dominance in confusion matrices and consistent per-class F1-scores across materials. These architectures effectively preserve fine thermal gradients and mitigate vanishing gradient effects, which is critical when distinguishing materials with overlapping emissivity ranges. Lightweight models such as MobileNet-V3 (86.4%) provide substantial computational advantages, achieving inference speeds up to 168 FPS with only 5.4 million parameters. However, reduced representational capacity leads to increased confusion among insulating materials such as fabric and XPS foam, where subtle cooling differences demand higher feature resolution. Interestingly, the CNN+LSTM hybrid architecture (84.7%) did not outperform optimized convolutional backbones. Although temporal modeling captures dynamic decay behavior, sequential processing may limit parallel spatial abstraction strength. This suggests that spatial thermal gradients contribute more strongly to separability than purely temporal embeddings under controlled acquisition conditions. The proposed HybridNet, integrating CNN, transformer, and bidirectional LSTM modules, achieved 82.3% in-distribution accuracy. While this does not surpass EfficientNet-B4, radar analysis indicates relatively improved robustness characteristics under distribution shifts. The adaptive cross-modal attention mechanism may therefore offer advantages in less controlled or field-deployed scenarios, where environmental perturbations degrade individual feature streams. A key observation across all models is that misclassification patterns align closely with thermophysical similarity rather than arbitrary visual resemblance. High-conductivity materials such as steel (16.2 $W \cdot m^{-1} \cdot K^{-1}$) and alumina (30.0 $W \cdot m^{-1} \cdot K^{-1}$) exhibit rapid cooling and strong separability. In contrast, ultralow-conductivity materials such as XPS foam (0.03 $W \cdot m^{-1} \cdot K^{-1}$) and polyester fabric (0.04 $W \cdot m^{-1} \cdot K^{-1}$) demonstrate prolonged thermal decay, increasing inter-class overlap. Mid-conductivity materials such as concrete (1.8 $W \cdot m^{-1} \cdot K^{-1}$), glass (1.0 $W \cdot m^{-1} \cdot K^{-1}$), and CFRP (5.0 $W \cdot m^{-1} \cdot K^{-1}$) form the primary confusion clusters. These results confirm that transient thermal imaging encodes physically meaningful information, and that classification difficulty scales with proximity in thermal conductivity and emissivity space. The strong correlation between thermophysical properties and confusion trends reinforces the interpretability of the learned feature representations. While the proposed framework demonstrates strong performance in controlled experimental conditions, several limitations should be acknowledged. First, the ThermalMat-85K dataset, although diverse in geometry and environment, is still acquired under partially controlled conditions, which may not fully represent real-world thermal variability. Second, the HybridNet architecture introduces increased computational complexity and exhibits lower in-distribution accuracy compared to optimized convolutional models, indicating that its advantages are not fully realized under limited variability conditions. Third, the temporal sequence length is relatively short, which may constrain the effectiveness of temporal modeling components. Finally, the reliance on ImageNet-pretrained initialization highlights the current lack of large-scale thermal-domain pretraining resources. Future work will focus on expanding dataset diversity, exploring longer temporal sequences, and developing domain-specific pretraining strategies to further enhance performance and generalization.

6. Conclusion

This study demonstrates that transient thermal infrared signatures provide a robust and physics-informed basis for intelligent material recognition. By constructing the ThermalMat-85K dataset under controlled multi-angle and multi-environment conditions and validating all materials through FT-IR and XRD characterization, we established a reliable foundation for benchmarking deep learning architectures in thermal classification tasks. Among the evaluated models, EfficientNet-B4 achieved the best overall performance, delivering 93.1% test accuracy with balanced precision and recall across all classes. Residual and dense architectures also exhibited strong generalization, while lightweight models offered real-time capability at reduced accuracy. Hybrid temporal architectures demonstrated potential robustness benefits but did not outperform optimized convolutional backbones under laboratory conditions. Importantly, class-level confusion analysis revealed that errors are governed by thermophysical proximity, confirming that heat diffusion behaviour serves as the primary discriminative factor. This physics-aligned behaviour enhances the interpretability and reliability of the proposed framework. Future work will focus on expanding environmental variability, integrating domain adaptation strategies, and exploring self-supervised thermal representation learning to further improve robustness in real-world deployments. The presented dataset and benchmarking study provide a comprehensive foundation for advancing intelligent, non-contact material recognition systems based on thermal infrared imaging.

References

[1] N. F. M. Taha, "Robotics as a catalyst for sustainable food manufacturing: A malaysian empirical study," *Applied Research*, vol. 5, p. e70077, 2026.

[2] C. Shaba, D. P. Elpa, and P. L. Urban, "Automated and robotic sample delivery systems for mass spectrometry and ion-mobility spectrometry," *Digital Discovery*, vol. 5, no. 3, pp. 983–1010, 2026.

[3] X. C. A. Chacon, G. Cappuccino, and R. A. Lu, "Smart sorting systems of spent lithium batteries: Challenges and pathways to industrial adoption," *Journal of Energy Storage*, vol. 153, p. 120971, 2026.

[4] Q. Zeng, H. Liu, Y. Zhang, Y. Zheng, X. Ding, M. Zhu, G. Shi,

Y. Wang, H. Haick, and M. Zhang, "Materials and system design for self-decision bioelectronic systems," *Advanced Materials*, vol. 38, no. 9, p. e21164, 2026.

[5] D. Li, H. Shi, B. Cai, X. Bai, L. Chen, G. Han, and C. Mi, "A review of technical advances and applications of intelligent inspection robots in structural health monitoring," *SmartBot* vol. 1, no. 3, p. e70000, 2025.

[6] S. Hajra, S. Panda, K. R. Kaja, M. A. Belal, V. Vivekananthan, and H. J. Kim, "Self-powered fire safety indicator based on fabric-based triboelectric nanogenerator," *Energy Technology*, vol. 13, no. 10, p. 2402488, 2025.

[7] K. R. Kaja, S. Hajra, S. Panda, M. Belal, S. Nam, P. Pakawanit, B. K. Panigrahi, H. Khanbareh, C. Bowen, J. Yu, and H. J. Kim, "Waste polyethylene-coated fabrics for dual-mode interfaces triboelectrification for self-powered sensors," *Results in Engineering*, vol. 28, p. 107111, 2025.

[8] J.-T. Lin, Y.-P. Chung, Y.-Y. Lee, T.-L. Wu, T. T. T. Huynh, P. T. Nguyen, M.-J. Lu, B.-W. Huang, B. Sriram, S.-F. Wang, G.-P. Chang-Chien, S. Kogularasu, and W.-C. Lin, "Emerging analytical frontiers in microplastic detection: From spectroscopy to smart sensor technologies," *Talanta Open*, vol. 12, p. 100514, 2025.

[9] A. Hernik, J. Smets, Z. Wang, Y. Semenova, J.-C. Tan, R. Ameloot, and I. Naydenova, "Optical detection of volatile organic compounds: A review of methods and functionalized sensing materials," *Advanced Optical Materials*, vol. 13, no. 27, p. e00369, 2025.

[10] M. Inês Silva, E. Malitckii, T. G. Santos, and P. Vilaça, "Review of conventional and advanced non-destructive testing techniques for detection and characterization of small-scale defects, *Progress in Materials Science*, vol. 138, p. 101155, 2023.

[11] M. Shaaban, Y. Al-Hamidi, S. El-Borgi, and A. Krishnamoorthy, "Machine learning-driven in situ defect monitoring and real-time process control in directed energy deposition: Techniques, challenges, and future prospects," *Materials Today Communications*, vol. 51, p. 114767, 2026.

[12] P. Salazar-Villacis, and B. Benyahia, "The ADePT framework for assessing autonomous laboratory robotics," *Communications Chemistry*, vol. 9, Art. no. 99, 2026.

[13] G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E. M. Rajaonson, M. Skreta, N. Yoshikawa, S. Corapi, G. D. Akkoc, F. Strieth-Kalthoff, M. Seifrid, and A. Aspuru-Guzik, "Self-driving laboratories for chemistry and materials science," *Chemical Reviews*, vol. 124, no. 16, 9633–9732, 2024.

[14] S. Hajra, K. R. Kaja, S. Panda, M. A. Belal, B. K. Panigrahi, P. Pakawanit, and H. J. Kim, "Waste based triboelectric nanogenerator for energy harvesting and self-powered sensors," *Journal of Cleaner Production*, vol. 509, Art. no.145591, 2025.

[15] S. Mekruksavanich, and A. Jitpattanakul, "Hybrid convolution neural network with channel attention mechanism for sensor-based human activity recognition," *Scientific Reports*, vol. 13, no. 1, p. 12067, 2023.

[16] P. Das, A. Barma, F. R. Rijvi, A. Islam, and M. R. Rashid, "Unified analytical framework for non-Fourier heat conduction in graded cylindrical media: Applications to perovskite solar cells and borehole heat exchangers," *Case Studies in Thermal Engineering*, vol. 79, p. 107818, 2026.

[17] P. Saini, J. D. Osorio, M. P. Shah, U. N. Patel, A. Nelapudi, and L. A. Porto-Hernandez, "Nature-inspired lotus-shaped fins combined with hybrid nanoparticles and metal foam for high-performance latent heat thermal energy storage," *International Journal of Heat and Mass Transfer*, vol. 255, p. 127897, 2026.

[18] B. Ozdemir, and I. Pacal, "An innovative deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self-attention mechanisms," *Results in Engineering*, vol. 25 p. 103692, 2025.

[19] X. Cui, Q. Ruan, X. Zhuo, X. Xia, J. Hu, R. Fu, Y. Li, J. Wang, and H. Xu, "Photothermal nanomaterials: A powerful light-to-heat converter," *Chemical Reviews*, vol. 123, no. 11, p. 6891–6952, 2023.

[20] A. Batool, and Y.-C. Byun, "A lightweight multi-path convolutional neural network architecture using optimal features selection for multiclass classification of brain tumor using magnetic resonance images," *Results in Engineering*, vol. 25, p. 104327, 2025.

[21] L. Gao, S. Zheng, and F. Min, "WAHNet: A wave-attention hybrid network for surface wave suppression in seismic data," *Geophysical Prospecting*, vol. 74, no. 3, p. e70151, 2026.

[22] S. D. Alpsalaz, E. Aslan, Y. Özüpak, F. Alpsalaz, H. Uzel, and V. Bereznychenko, "Hybrid deep learning with attention fusion for enhanced colon cancer detection," *Scientific Reports*, vol. 15, Art. no. 45583, 2025.

[23] S. Oh, E. Lee, and S. K. Hong, "Physic-informed deep operator networks for modeling 2D time-domain electromagnetic wave propagation in various media," *IScience*, vol. 29, no. 3, p. 115076, 2026.

[24] G. Li, Q. Fang, L. Zha, X. Gao, and N. Zheng, "HAM: Hybrid attention module in deep convolutional neural networks for image classification," *Pattern Recognition*, vol. 129, p. 108785, 2022.

[25] M. Z. Naser, "Fundamental flaws of physics-informed neural networks and explainability methods in engineering systems," *Computers & Industrial Engineering*, vol. 212, p. 111704, 2026.

[26] G. Subbiah, M. Channappagoudra, D. Bhanot, V. J. Upadhye, S. Choudhury, K. K. Shukla, and K. P. Kumar, "Bio-based innovation: Assessing the suitability of Solanum nigrum stem fibers for textile and biomedical applications," *Results in Engineering*, vol. 26, p. 105441, 2025.

[27] A. Otsuki, S. Jose, M. Mohan, and S. Thomas, "Chapter 6 - Scanning probe microscope" in *Non-Destructive Material Characterization Methods*, Elsevier, 2023, pp. 151–165.

[28] J. George, K. Palanisamy, S. Kulandaivel, P. Saravanan, D. M. Peedika, K. Rajagopalan, and M. K. Kaliyannan, "Impact of cerium oxide nanoparticles on survivability and reproduction of earthworm eudrilus eugeniae and its compost quality," *BioNanoScience*, vol. 13, no. 2, pp. 1911–1921, 2023.

[29] S. M. Hosseini-Nasab, S. H. R. Mousavi, and S. Fuchs, "Thermal-property profiles from well-logs in sedimentary rocks: A novel machine-learning-based prediction tool trained on physically modelled synthetic data," *Geophysical Journal International*, vol. 243, no. 1, p. ggaf260, 2025.

[30] N. Karimi, F. Riffard, F. Rabaste, S. Perrier, R. Cueff, C. Issartel, and H. Buscail, "Characterization of the oxides formed at 1000°C on the AISI 304 stainless steel by X-ray diffraction and infrared spectroscopy," *Applied Surface Science*, vol. 254, no. 8, pp. 2292–2299, 2008.

[31] Y.-J. Cai, T.-Y. Liu, T. Shen, Y.-H. Wu, Z.-Z. Lin, G.-Y. Zhu, and C. Wang, "The crystallization properties of PVC were measured by room-temperature phosphorescence," *Materials Today Communications*, vol, 47, p. 113044, 2025.

[32] K. Djebaili, Z. Mekhalif, A. Boumaza, and A. Djelloul, "XPS, FTIR, EDX, and XRD analysis of Al_2O_3 scales grown on PM2000 alloy," *Journal of Spectroscopy*, vol. 2015, no. 1, p. 868109, 2015.

[33] T. Suter, V. Brázdová, K. McColl, T.S. Miller, H. Nagashima, E. Salvadori, A. Sella, C. A. Howard, C. W. M. Kay, F. Corà, and P. F. McMillan, "Synthesis, structure and electronic properties of graphitic carbon nitride films," *The Journal of Physical Chemistry C*, vol. 122, no. 44, pp. 25183–25194, 2018.

[34] M. Bosca, L. Pop, G. Borodi, P. Pascuta, and E. Culea, "XRD and FTIR structural investigations of erbium-doped bismuth–lead–silver glasses and glass ceramics," *Journal of Alloys and Compounds*, vol. 479, no. 1–2, pp. 579–582, 2009.

[35] Y. J. Lee, S. W. Kweon, T.-J. Lee, and H. J. Kim, "Quantification of cellulose II using infrared spectroscopy: Machine learning approaches," *International Journal of Biological Macromolecules*, vol. 323, Art. no. 147193, 2025.