Ilmiy tadqiqotlar va ularning yechimlari jurnali
Aoka guvohnoma № 323759 ISSN: 3060-4818

A REVIEW OF MACHINE LEARNING AND DEEP LEARNING APPLICATIONS IN HEALTHCARE, AFFECTIVE COMPUTING, SAFETY MONITORING, AGRICULTURE, CYBERSECURITY AND EDUCATION

Alpamis Kutlimuratov
Amirxon Maxsudxonov
Department of Applied Informatics,
Kimyo International University
in Tashkent, Tashkent 100121, Uzbekistan
kutlimuratov.aj@kiut.uz, a.maxsudxonov@kiut.uz
https://doi.org/10.5281/zenodo.24393467

Abstract. Machine learning (ML) and deep learning (DL) have moved from benchmark studies to systems that support clinical diagnosis, public safety, food production, network protection and education. This review summarises recent progress in these application areas and identifies the methodological patterns they share. We first outline the core model families used in applied work, including handcrafted features with ensemble classifiers, convolutional neural networks, Transformers and hybrid architectures, lightweight models for edge devices, multimodal fusion and generative and large language models. We then examine seven application domains: medical image analysis, clinical text simplification, affective computing and speech security, fire and aerial monitoring with unmanned aerial vehicles, plant disease recognition and agricultural decision support, network intrusion and anomaly detection, and sequence-aware recommendation in e-learning. For each domain we discuss typical tasks, datasets and model choices, and we summarise representative studies in comparison tables. Across domains, five trends stand out. Models are becoming lighter and are increasingly evaluated on edge hardware. Data scarcity is addressed with augmentation, transfer learning and generative models. Multimodal and multi-stream fusion is widely used to improve robustness. Robustness to domain shift is becoming a central evaluation goal, and growing attention is paid to data leakage, reproducibility and explainability. We conclude with open challenges, including standardised efficiency reporting, cross-dataset evaluation, field validation and the responsible use of generative models, and outline directions for future research.

Keywords: machine learning; deep learning; medical image analysis; affective computing; fire detection; plant disease recognition; intrusion detection; lightweight models; multimodal fusion; explainable AI

1. Introduction

Deep learning has changed how computers analyse images, speech, text and signals [1]. Architectures such as residual networks [2], U-Net [3] and Transformers [4,5] now form the basis of applied systems in almost every sector. In parallel, classical ML methods such as gradient-boosted trees [6] remain competitive when data are limited or when features can be designed from domain knowledge. As a result, ML and DL are no longer confined to computer science laboratories. They are used to segment tumours in magnetic resonance images, to recognise emotions from faces and voices, to detect forest fires from drones, to diagnose plant diseases from leaf photographs and to stop attacks on computer networks.

Most existing reviews focus on a single domain or on a single family of methods, such as lightweight networks [7,8], tiny machine learning [9], multimodal fusion [10,11] or trustworthy and explainable AI [12,13]. Such reviews are valuable, but they make it difficult to see which ideas transfer between domains. A technique that improves small-object detection in aerial images may also help with small lesions in medical images, and an evaluation lesson learned in one field may prevent an error in another. Applied researchers, especially in smaller research groups that work across several sectors, need an overview that connects these fields.

This review addresses that need. It covers recent applied ML and DL work in seven domains and draws out the methodological patterns they share. The contributions are as follows. First, we give a compact overview of the model families that dominate applied work. Second, we review applications in medical image analysis, clinical natural language processing, affective computing and speech security, fire and aerial monitoring, agriculture, cybersecurity and education, and summarise representative studies in comparison tables. Third, we identify five cross-domain trends and discuss open challenges and future directions.

The rest of the paper is organised as follows. Section 2 introduces the core methods. Section 3 describes how the literature was selected. Section 4 reviews the application domains. Section 5 discusses cross-domain trends, Section 6 presents challenges and future directions, and Section 7 concludes.

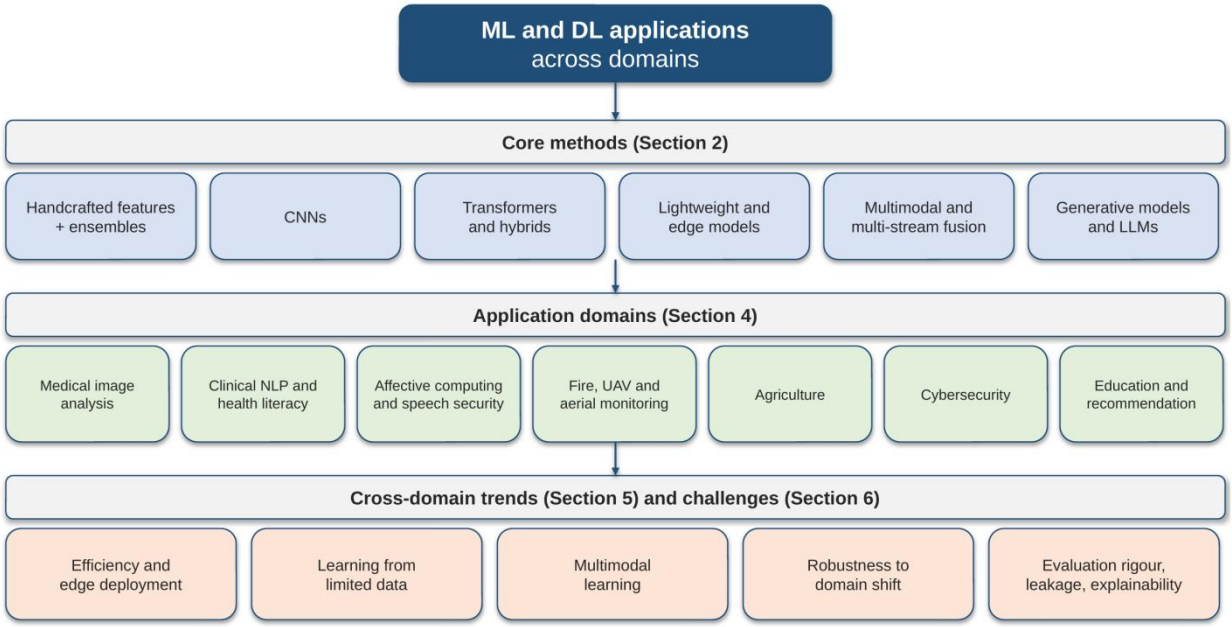

Figure 1. Organisation of the review. Core method families (Section 2) are applied in seven domains (Section 4), which share common trends and challenges (Sections 5 and 6).

2. Overview of Core Methods

2.1. Handcrafted Features and Ensemble Classifiers

Before deep learning, image and signal analysis relied on features designed by experts. Texture descriptors such as grey-level co-occurrence statistics [14] and local binary patterns [15] capture local intensity structure and remain useful when training data are scarce. Fractal measures describe the self-similarity of irregular patterns such as diseased plant tissue [16]. These features are usually combined with classifiers such as support vector machines or gradient-boosted trees. XGBoost [6], in particular, is

widely used for tabular data and for small image datasets after feature extraction [17]. Handcrafted features are also combined with deep networks, for example by feeding Gabor and local binary pattern maps into a convolutional network for facial emotion recognition [18].

2.2. Convolutional Neural Networks

Convolutional neural networks (CNNs) learn hierarchical image features directly from pixels [1]. Residual connections made very deep networks trainable [2], and encoder–decoder designs such as U-Net [3] became the standard for medical image segmentation. One-stage detectors such as YOLO [19] made real-time object detection practical and are now the most common choice for fire, smoke and aerial monitoring [20,21]. Compound scaling, as in EfficientNet [22], and depthwise separable convolutions, as in MobileNets [23], made CNNs efficient enough for embedded devices.

2.3. Transformers and Hybrid Architectures

The Transformer replaced recurrence with self-attention [4] and was first used in natural language processing, where pre-trained models such as BERT [24] became the default starting point. The Vision Transformer [5] and the Swin Transformer [25] brought the same idea to images. In applied work, pure Transformers are often combined with convolutional layers, because convolutions provide a useful bias for local structure while attention captures long-range context. Hybrid designs such as TransUNet [26] and Swin UNETR [27] are now common in medical image segmentation [28], and hybrid Transformer–CNN models with boundary-aware attention have been proposed for multi-modal brain tumour segmentation [29]. Landmark-guided Transformers have been used for facial emotion recognition [30], and graph neural networks [31] are used when the data have an explicit structure, such as facial action units [32].

2.4. Lightweight and Edge-Oriented Models

Many applications require inference on drones, cameras, smartphones or network appliances rather than on servers. Lightweight model design, pruning, quantisation and knowledge distillation [33] are therefore central topics [7,34]. Recent surveys review lightweight object detectors for edge devices [8] and tiny machine learning on microcontrollers [9]. Applied studies increasingly report not only accuracy but also model size and speed, for example in lightweight UAV-based fire-risk identification [35], cascade-based zero-day attack detection [36] and dynamic facial expression recognition on edge devices using compressed-domain temporal cues and degradation-aware distillation [37].

2.5. Multimodal and Multi-Stream Fusion

Many real problems involve more than one source of information, such as speech and facial video, several MRI sequences, or images and sensor readings. Multimodal fusion combines these sources at the input, feature or decision level [10,11]. In emotion recognition, fusion of audio, visual and textual cues is now standard [38], and dual-branch attention with hierarchical adaptive fusion has been proposed to weight modalities dynamically [39]. Multi-stream designs are also used within a single modality, for example by processing prosodic and spectral streams separately to detect AI-synthesised speech [40], or by combining several views of a scene for human–object interaction detection [41,42].

2.6. Generative Models and Large Language Models

Generative adversarial networks [43] and diffusion models [44] can create realistic synthetic images and are used to augment small datasets, particularly in medicine [45,46]. Large language models (LLMs) encode broad knowledge and can answer clinical questions [47], but they also produce fluent errors. Retrieval-augmented generation (RAG) [48] grounds the output of an LLM in trusted documents and has been applied to clinical medicine [49]. Generative models are also a source of new threats, because synthetic speech and images can be used for fraud, which in turn motivates detection research [40,50].

2.7. Evaluation Metrics

Each domain has its own standard metrics, which makes direct comparison between domains difficult. Table 4 lists the most common choices. Classification tasks report accuracy, precision, recall, F1-score and the area under the ROC curve. Segmentation tasks use the Dice coefficient and the Hausdorff distance [51]. Detection tasks use mean average precision at one or more intersection-over-union thresholds [19,52]. Spoofing detection uses the equal error rate and the tandem detection cost function [53], and intrusion detection pays particular attention to the false positive rate, because a high false alarm rate makes a system unusable in practice [54]. For deployment, an increasing number of studies also report the number of parameters, floating-point operations, model size and inference latency [7]. In each case, the metric should match the cost of errors in the application. Missing a tumour, a fire or an attack is usually more costly than a false alarm, so recall at a fixed false alarm rate is often more informative than overall accuracy.

Table 4. Common evaluation metrics by domain.

Domain	Typical task	Main metrics	Efficiency metrics	Ref.
Medical imaging	Segmentation, classification	Dice, Hausdorff distance, AUC, sensitivity	Parameters, inference time per volume	[27,51]
Clinical NLP	Text simplification	Readability indices, semantic similarity, expert rating	Generation latency	[47,55]
Affective computing	FER, SER	Accuracy, unweighted average recall, F1	FLOPs, frames per second	[56,57]
Speech security	Spoofing detection	Equal error rate, t-DCF	Real-time factor	[53]
Fire and aerial monitoring	Object detection	mAP@0.5, mAP@0.5:0.95, false alarm rate	FPS on edge device, model size	[52,58]
Agriculture	Disease classification	Accuracy, F1, confusion matrix	Model size, mobile latency	[59]
Cybersecurity	Intrusion detection	Detection rate, false positive rate, F1	Throughput, latency	[54,60]

3. Review Approach

This is a narrative review. Relevant studies were identified through searches of Scopus, Web of Science, IEEE Xplore, PubMed and Google Scholar using combinations of domain terms (for example, brain tumour segmentation, speech emotion recognition, forest fire detection, plant disease, intrusion detection and e-learning recommendation) with method terms (machine learning, deep learning, convolutional neural network, Transformer, lightweight, multimodal). We focused on peer-reviewed journal and conference papers published mainly between 2022 and 2026, and added earlier landmark papers where they introduced methods or datasets that are still in use. Recent surveys in each domain were used to identify further studies. We included studies that proposed or evaluated an ML or DL method on a named dataset and reported quantitative results. Studies were grouped by application domain and then compared by task, data modality and model design. Because metrics and datasets differ between domains, we did not attempt a quantitative meta-analysis.

4. Applications by Domain

4.1. Medical Image Analysis

Medical image analysis is one of the most active application areas of DL [28]. Brain tumour segmentation from multi-sequence MRI is a long-standing benchmark task, supported by the BraTS challenge datasets [51]. Hybrid models that combine convolution and attention, such as Swin UNETR [27] and TransUNet [26], have become strong baselines, and further work adds boundary-aware attention to sharpen tumour edges across MRI modalities [29]. Beyond segmentation, multi-task models now predict several clinical targets from the same image, linking anatomical appearance to molecular and genomic markers of glioma [61].

Microscopy is a second important area. Blood smear images are used to diagnose anaemia, leukaemia and other blood disorders, and image quality varies with staining and acquisition. Studies have shown that careful image preprocessing combined with deep classification models improves early diagnosis of blood disorders [62]. A common problem across medical imaging is the small size of labelled datasets. Classical augmentation, generative models and efficient augmentation pipelines are therefore widely used [45], including methods designed specifically to enhance medical images during augmentation [46]. Multimodal approaches that combine images with clinical records are also growing [11].

Several practical issues limit the clinical use of these models. Datasets are often collected at a single hospital with a single scanner, so performance may drop when the model is applied elsewhere. Labels for segmentation require time from expert radiologists, and inter-observer variation adds noise to the ground truth. Patient-level data splits are essential, because several images or slices from the same patient in both training and test sets inflate the reported accuracy [63]. Explainability tools such as Grad-CAM [64] are commonly used to show which image regions drive a prediction, but clinicians also need calibrated confidence and clear information on when the model should not be trusted. Multi-task designs that predict anatomical and molecular targets together [61] and boundary-aware segmentation [29] are steps towards outputs that are more directly useful in clinical workflows.

4.2. Clinical Natural Language Processing and Health Literacy

Medical documents and instructions are often too complex for patients, especially older adults. Text simplification aims to rewrite such content in plain language without changing its meaning. LLMs make

this feasible at scale [47], but they may introduce errors or omit important warnings. Recent work therefore adds domain control to the generation process. Examples include a generative expert-narrated simplification model designed to improve health literacy among older people [65], a domain-aware RAG system that simplifies rehabilitation instructions using trusted sources [66], and a neuro-semantic clinical filter that checks simplified medical text for semantic consistency [55]. These approaches reflect a wider move towards retrieval-grounded clinical language models [48,49].

Evaluation is a central difficulty in this area. Automatic readability scores measure how simple a text is but not whether it is still correct, and similarity metrics reward copying rather than simplification. For this reason, studies increasingly combine automatic measures with expert review of factual accuracy and with checks for omitted safety information [55]. Transformer encoders such as BERT [24] remain useful as components for semantic checking, while generation is delegated to larger models [47]. Privacy is a further constraint, because clinical texts contain personal data, which favours smaller models that can run inside the hospital network.

4.3. Affective Computing and Speech Security

Affective computing aims to recognise human emotions from faces, voices and other signals. Facial expression recognition (FER) is commonly evaluated on datasets such as FER2013 [57], while speech emotion recognition (SER) uses acted and spontaneous corpora such as IEMOCAP [67]. Reviews show a shift from handcrafted acoustic features to deep spectral representations and Transformers [56]. In SER, work has focused on high-level data representations and fast feature extraction to reduce computation [68]. In FER, several directions coexist. Handcrafted texture filters can be combined with CNNs [18]. Facial landmarks can guide attention in Transformers [30]. Facial action units can constrain graph-based models, and counterfactual explanations [69] can make predictions auditable [32]. Temporal cues in compressed video can support lightweight dynamic FER on edge devices [37]. Multimodal emotion recognition combines these signals, and adaptive fusion helps when one modality is noisy or missing [38,39]. Face detection and recognition, which precede many of these pipelines, have been reviewed in detail [70].

A related problem is the detection of synthetic speech. Text-to-speech and voice conversion systems can now imitate a speaker convincingly, which creates risks for voice authentication and public trust. The ASVspoof challenges provide standard datasets for spoofing detection [53], and recent surveys summarise the rapid growth of this field [50]. A key difficulty is generalisation to unseen generators and to audio that has passed through different codecs. Prosody-guided multi-stream frameworks have been proposed to detect AI-synthesised speech across codec and vocoder domains [40].

Affective computing also raises questions of fairness and validity. Datasets differ in whether emotions are acted or spontaneous, in recording conditions and in the demographic mix of participants, so a model trained on one corpus may not transfer to another [56]. Labels are subjective, and agreement between annotators is often moderate. For these reasons, recent work emphasises models whose decisions can be traced to interpretable cues, such as facial action units [32] or facial landmarks [30], and lightweight models that can run on the user's device and keep video and audio private [37].

4.4. Fire Safety, UAVs and Aerial Monitoring

Early detection of fire and smoke saves lives and reduces damage to forests, buildings and ships. Vision-based detectors, mostly variants of YOLO [19], are trained on datasets such as D-Fire [58] and are

reviewed in several recent surveys [71,72]. Unmanned aerial vehicles (UAVs) extend the coverage of fixed cameras and allow early warning in remote forests [71,73]. Studies have proposed hybrid DL models for early forest fire detection [21], lightweight UAV-based systems for early fire-risk identification [35] and detectors adapted to the specific conditions of maritime fire safety [20]. Evaluation practice is an important issue in this field, because fire datasets often contain near-duplicate frames from the same video. A leakage-aware evaluation framework and benchmark for edge-based multi-sensor fire and smoke detection has shown how such duplicates can inflate reported accuracy [74].

Vision is not the only signal available for fire detection. Smoke, temperature and gas sensors respond to fires that are not yet visible, while cameras cover large areas and localise the fire. Multi-sensor systems that combine these sources can reduce false alarms caused by fog, sunlight or steam, which are common failure cases of image-only detectors [72,74]. Deployment conditions also differ strongly between applications. Forest monitoring relies on UAVs or towers with limited power and connectivity [35,73], maritime fire safety must cope with reflections, waves and confined ship interiors [20], and urban systems must process many camera streams at once. These conditions explain why lightweight detectors and careful evaluation on realistic data are central themes in this field [8].

Aerial imagery raises general challenges beyond fire. Objects such as people and vehicles may occupy only a few pixels, which makes detection difficult [52,75]. Measure-based supervision has been proposed to improve tiny object detection in aerial images [76]. UAVs are also used in urban mobility and traffic monitoring, where the identity of participating devices must be validated, and context-aware identity validation has been proposed for UAV-assisted urban environments [77].

4.5. Agriculture

Plant diseases cause major crop losses, and image-based diagnosis can help farmers act early. Public datasets such as PlantVillage [78] enabled rapid progress, and recent systematic reviews summarise the growing use of CNNs, Transformers and lightweight models for plant disease recognition [59,79]. Deep models dominate, but classical pipelines remain competitive on small datasets. For example, texture descriptors combined with an XGBoost ensemble have been used to classify tomato leaf diseases [17], and fractal-based image analysis has been applied to the multi-stage detection of tomato late blight in greenhouse-grown plants [16]. ML also supports decisions beyond the field. In agribusiness, a hybrid of the NSGA-II multi-objective algorithm [80] and ML models has been used to optimise marketing budgets in Kazakhstan's agro-industrial complex [81].

A known weakness of the field is the gap between laboratory images and field conditions. Many datasets, including PlantVillage [78], contain single leaves photographed against a plain background, while field images contain several leaves, soil, shadows and variable lighting. Models that reach very high accuracy in the laboratory often lose much of it in the field [59,79]. Lightweight models that run on smartphones or edge devices are therefore an active topic [8], together with datasets collected under real growing conditions, such as greenhouse images for multi-stage disease detection [16].

4.6. Cybersecurity

Intrusion detection systems (IDS) monitor network traffic to identify attacks [54]. Benchmark datasets such as NSL-KDD [60] and CICIDS2017 [82] support the comparison of ML and DL methods, and recent work explores advanced learning paradigms for intelligent networks [83]. A persistent

challenge is the detection of zero-day attacks that are absent from the training data [84]. Lightweight cascade-based frameworks aim to detect such attacks in real time with low computational cost [36]. Software-defined networking (SDN) centralises network control and creates new opportunities for detection and response, and hybrid approaches that integrate anomaly detection with automated mitigation have been proposed for SDN environments [85].

Benchmark datasets in this field have well-known limitations. Older datasets do not reflect modern traffic, and newer ones contain duplicated flows and strong class imbalance, which can make simple classifiers look very accurate [54,82]. Evaluation on attacks that are absent from the training set [84] and on traffic from different networks gives a more realistic picture. In practice, detectors must also run at line speed on limited hardware, which explains the interest in lightweight and cascade designs that pass only suspicious traffic to more expensive models [36].

4.7. Education and Recommender Systems

In e-learning, recommender systems help learners find relevant courses and materials. Learner activity forms a time-ordered sequence, so sequence models such as recurrent networks [86,87] and deep candidate-generation models [88] are well suited to the task. A deep sequence-aware candidate generation method has been proposed for e-learning systems, which models the order of learner interactions to improve recommendations [89]. Similar sequential modelling ideas appear in other domains, for example in temporal facial expression analysis [37], which illustrates how methods move between fields.

Educational recommendation also raises issues that are common to other human-centred applications. Learner data are personal, so privacy must be protected. Recommendations should be explainable to teachers and students, and evaluation should consider not only click or completion rates but also learning outcomes. Offline accuracy on historical logs may not reflect the effect of a recommender on real learners, which mirrors the gap between benchmark and field performance discussed for other domains in this review.

4.8. Human-Centred Visual Understanding

Several of the domains above require the system to understand what people are doing, not only what objects are present. Human–object interaction (HOI) detection identifies triplets of person, action and object in an image, and is evaluated on benchmarks such as HICO-DET [41]. HOI detection is relevant to safety monitoring, for example to recognise dangerous behaviour near fire sources, and to assistive technologies in healthcare. Multi-view context learning has been proposed to improve HOI detection by combining information from several views of a scene [42]. This line of work links object detection, pose understanding and affective computing, and it illustrates how general computer vision advances feed into the applied domains reviewed here.

Table 1. Summary of representative application studies by domain.

Domain	Typical tasks	Typical data	Main model families	Representative studies
Medical image analysis	Tumour segmentation, multi-task glioma profiling, blood disorder diagnosis, augmentation	MRI, microscopy	Hybrid CNN–Transformer, multi-task CNN, generative augmentation	[26,27,29,45,46,61,62]
Clinical NLP	Medical text simplification, patient instruction generation	Clinical and patient texts	LLMs, RAG, semantic filters	[47,49,55,65,66]
Affective computing	FER, SER, multimodal emotion recognition	Face images and video, speech	CNN, Transformer, graph models, multimodal fusion	[18,30,32,37–39,56,68]
Speech security	Synthetic speech detection	Speech audio	Multi-stream networks, prosodic features	[40,50,53]
Fire and aerial monitoring	Fire and smoke detection, tiny object detection, UAV identity validation	Camera and UAV images, sensor data	YOLO-family detectors, lightweight CNN	[20,21,35,58,73–77]
Agriculture	Leaf disease classification, disease stage detection, budget optimisation	Leaf images, tabular data	Texture features with XGBoost, fractal analysis, NSGA-II with ML	[16,17,59,78,81]
Cybersecurity	Intrusion and zero-day attack detection, SDN anomaly mitigation	Network traffic	Ensembles, cascades, hybrid detectors	[36,82–85]
Education	Course and content recommendation	Learner interaction sequences	Sequence-aware deep models	[87–89]

Table 2. Widely used public benchmark datasets in the reviewed domains.

Domain	Dataset	Data type	Main task	Ref.
Medical imaging	BraTS	Multi-sequence brain MRI	Brain tumour segmentation	[51]
Affective computing	FER2013	Face images	Facial expression recognition	[57]

Domain	Dataset	Data type	Main task	Ref.
Affective computing	IEMOCAP	Audio, video, text	Speech and multimodal emotion recognition	[67]
Speech security	ASVspoof 2019	Speech audio	Spoofing and synthetic speech detection	[53]
Fire monitoring	D-Fire	Camera images	Fire and smoke detection	[58]
Aerial monitoring	VisDrone	UAV images and video	Object detection and tracking	[52]
Aerial monitoring	AI-TOD	Aerial images	Tiny object detection	[75]
Agriculture	PlantVillage	Leaf images	Plant disease classification	[78]
Cybersecurity	NSL-KDD	Network connection records	Intrusion detection	[60]
Cybersecurity	CICIDS2017	Network flows	Intrusion detection	[82]
Computer vision	HICO-DET	Images	Human–object interaction detection	[41]

5. Cross-Domain Trends

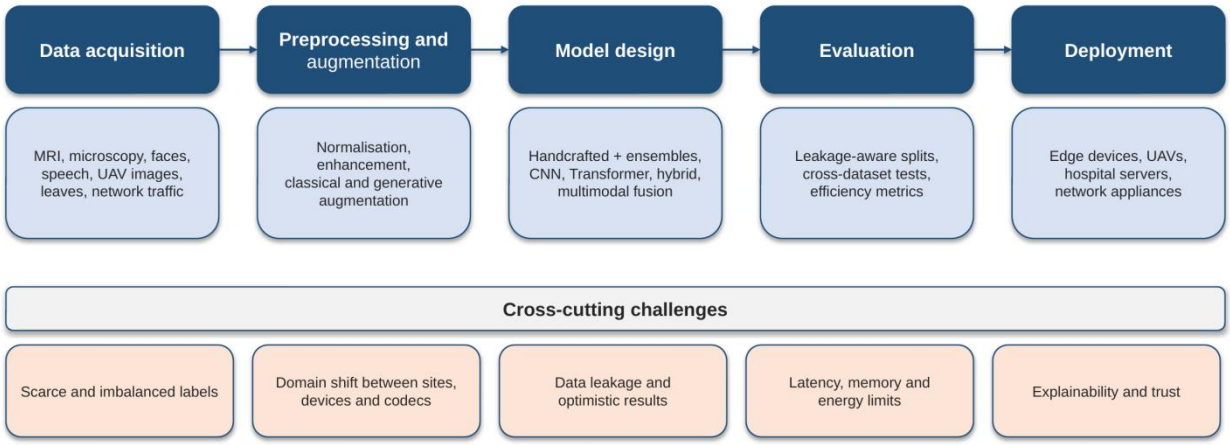

Figure 2. A generic pipeline for applied ML and DL systems, from data acquisition to deployment, with the cross-cutting challenges that appear in all reviewed domains.

Figure 2 shows the generic pipeline that underlies most of the studies reviewed above. Although the data and tasks differ, every application passes through the same stages of data acquisition, preprocessing, model design, evaluation and deployment, and faces similar challenges at each stage. The following subsections discuss the main trends that emerge when the domains are viewed together.

5.1. From Accuracy to Efficiency

A clear trend across domains is that efficiency is now reported alongside accuracy. Drones, cameras and network appliances have limited memory and power, and real-time response is often essential. Lightweight architectures [22,23], distillation [33] and compression [7,34] are applied in fire detection [35], cybersecurity [36] and affective computing [37]. However, efficiency is still reported inconsistently. Some studies give the number of parameters, others give floating-point operations, and only a minority report latency on named hardware [8]. Standardised efficiency reporting would make results easier to compare.

Efficiency gains also come from the system design, not only from the network. Cascade architectures run a cheap model on all inputs and pass only uncertain cases to a more expensive model [36]. Compressed-domain processing avoids full decoding of video [37]. Edge–cloud splitting keeps simple decisions on the device and sends difficult cases to a server. These strategies are often more effective than compressing a single large network, and they deserve more attention in comparative studies [9,34].

5.2. Coping with Limited Data

Labelled data are scarce in most applied domains, because annotation requires experts such as radiologists, plant pathologists or security analysts. Common responses include transfer learning from large pre-trained models [5,24], data augmentation [45], generative synthesis [43,44] and handcrafted features that encode prior knowledge [14,15]. The same toolkit appears in medical imaging [46,62], agriculture [16,17] and FER [18], which suggests that lessons on augmentation can be shared between these communities.

The use of synthetic data requires care. Generated images may reproduce the biases of the training set or contain artefacts that a classifier learns to exploit, and synthetic samples should never appear in the test set. Augmentation must also be applied after the data have been split, otherwise transformed copies of the same image end up in both training and test sets, which is a common form of leakage [63]. Reporting the exact augmentation pipeline and the split procedure is therefore as important as reporting the model architecture.

5.3. Multimodal and Multi-Stream Learning

Combining several sources of information improves robustness when one source is noisy or missing [10]. In healthcare, fusion of imaging and clinical data is growing [11]. In affective computing, audio, visual and textual cues are fused [38,39]. In security, separate streams capture complementary artefacts of synthetic speech [40]. The main open questions concern when to fuse, how to handle missing modalities, and how to keep fusion models efficient.

Fusion can happen at different levels. Early fusion concatenates raw inputs or low-level features, intermediate fusion combines learned representations, often with attention, and late fusion combines the decisions of separate models [10]. Intermediate fusion with attention has become the most common choice in recent work because it lets the model weight each modality according to its reliability [38,39]. Late fusion remains attractive when modalities are recorded by different devices or at different times, as in multi-sensor fire monitoring [74], because each branch can be trained and updated independently.

5.4. Robustness and Generalisation

Models that perform well on one dataset often fail on data from another hospital, camera, codec or network. This domain shift appears in every field reviewed here. Spoofing detectors must generalise to unseen generators and codecs [40,50]. Intrusion detectors must handle unseen attacks [36,84]. Fire and plant disease models must cope with new environments and lighting [72,79]. Cross-dataset evaluation is therefore becoming as important as in-distribution accuracy, although it is still reported in only part of the literature.

Several practical steps improve robustness. Training data can be collected from several sites or devices, and models can be tested on a held-out site rather than a random split. Augmentations that simulate the expected shift, such as codec compression for speech [40], weather effects for aerial images [73] or staining variation for microscopy [62], prepare the model for new conditions. Domain adaptation and test-time adaptation adjust the model with unlabeled data from the new domain. Finally, models should report calibrated confidence, so that inputs that differ from the training data can be flagged for human review rather than silently misclassified [12].

5.5. Evaluation Rigour, Reproducibility and Explainability

Data leakage, such as near-duplicate images in training and test sets or information about the label hidden in the features, can make results look much better than they are. Kapoor and Narayanan found that leakage affected at least 294 studies across 17 scientific fields [63]. Similar risks exist in applied DL, for example when consecutive video frames are split randomly, and leakage-aware protocols have been proposed for fire detection [74]. Explainability is a related concern. Saliency methods such as Grad-CAM [64] and feature attribution methods such as SHAP [90] are widely used, but they do not always reveal why a model fails. Structured approaches, such as constraining models with domain knowledge and providing counterfactual explanations [32,69], aim to make decisions auditable. These ideas are part of the broader agenda of trustworthy AI [12,13].

5.6. Transferable Lessons

Table 3 summarises techniques that were developed or refined in one domain and that are likely to be useful in others. The aim is not to claim that each transfer has already been demonstrated, but to point researchers to ideas that address similar problems in neighbouring fields.

Table 3. Techniques that may transfer between domains.

Technique	Where it is used	Problem it addresses	Possible use elsewhere	Ref.
Tiny object supervision	Aerial monitoring	Objects of a few pixels	Small lesions in medical images, early disease spots on leaves	[75,76]
Leakage-aware splitting	Fire detection	Near-duplicate frames inflate accuracy	Video-based FER, medical slices, network flows	[63,74]

Technique	Where it is used	Problem it addresses	Possible use elsewhere	Ref.
Multi-stream fusion	Speech security, emotion recognition	Complementary cues, missing modalities	Multi-sensor fire detection, multimodal medical diagnosis	[10,39,40]
Texture and fractal features with ensembles	Agriculture, FER	Small datasets	Microscopy, industrial inspection	[15–18]
Cascade detection	Cybersecurity	Real-time processing on limited hardware	Video surveillance, UAV monitoring	[36]
Knowledge-constrained and counterfactual explanations	FER	Auditable decisions	Clinical decision support, intrusion alerts	[32,69]
Retrieval grounding	Clinical NLP	Factual errors of generative models	Educational assistants, security reports	[48,49,66]

6. Challenges and Future Directions

Several challenges remain open across the reviewed domains. First, efficiency reporting should be standardised. Studies should state the hardware, numerical precision, batch size and latency, in addition to the number of parameters and operations [7,8]. Second, cross-dataset and cross-site evaluation should become routine, because in-distribution accuracy often overestimates real performance [63]. Third, field validation remains rare. Many fire, agricultural and medical systems are tested only on curated datasets, and prospective tests in real conditions are needed before deployment.

Fourth, generative models create both opportunities and risks. They can enrich small datasets [44,45] and simplify medical information for patients [55,65,66], but they can also produce errors and enable fraud through synthetic speech and images [40,50]. Methods for grounding, filtering and detecting generated content will therefore remain important. Fifth, explainability should move from post hoc visualisation towards models whose decisions can be checked against domain knowledge [13,32]. Finally, many applications concern regions with limited computing resources and limited local data, including Central Asia. Lightweight models, transfer learning and open benchmarks that include data from such regions would make the benefits of ML and DL more widely available [81].

Finally, the gap between research and practice deserves more attention. Few of the reviewed studies involve end users such as clinicians, firefighters, farmers or network administrators in the design or evaluation of the system. Usability, integration with existing workflows, maintenance after deployment and the cost of errors in a specific setting are rarely reported. Closer collaboration between ML researchers and domain experts, together with open release of code, trained models and data splits [63], would make results easier to reproduce and more likely to reach practical use. Education also has a role here, since a new generation of engineers must learn not only how to train models but also how to evaluate them honestly and deploy them responsibly.

7. Conclusions

This review has surveyed recent applications of ML and DL in medical image analysis, clinical language processing, affective computing and speech security, fire and aerial monitoring, agriculture, cybersecurity and education. Although these domains differ in data and goals, they face common problems. They need efficient models for real-time use, ways to learn from limited data, robust fusion of several information sources, generalisation to new conditions, and trustworthy evaluation and explanation. Progress on these shared problems in one domain often benefits the others. We hope that this cross-domain view will help researchers choose suitable methods, avoid common evaluation pitfalls and identify promising directions for future work.

References

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
2. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
3. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. Med. Image Comput. Comput.-Assist. Interv. (MICCAI), in Lecture Notes in Computer Science, vol. 9351, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
4. A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 5998–6008. [Online]. Available: arXiv:1706.03762
5. A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021. [Online]. Available: arXiv:2010.11929
6. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (KDD), 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
7. H.-I. Liu et al., "Lightweight deep learning for resource-constrained environments: A survey," ACM Comput. Surv., vol. 56, no. 10, pp. 1–42, 2024, doi: 10.1145/3657282.
8. P. Mittal, "A comprehensive survey of deep learning-based lightweight object detection models for edge devices," Artif. Intell. Rev., vol. 57, no. 9, Art. no. 242, 2024, doi: 10.1007/s10462-024-10877-1.
9. Y. Abadade, A. Temouden, H. Bamoumen, N. Benamar, Y. Chtouki, and A. S. Hafid, "A comprehensive survey on TinyML," IEEE Access, vol. 11, pp. 96892–96922, 2023, doi: 10.1109/ACCESS.2023.3294111.
10. K. Bayoudh, "A survey of multimodal hybrid deep learning for computer vision: Architectures, applications, trends, and challenges," Inf. Fusion, vol. 105, Art. no. 102217, 2024, doi: 10.1016/j.inffus.2023.102217.
11. F. Krones, U. Marikkar, G. Parsons, A. Szmul, and A. Mahdi, "Review of multimodal machine learning approaches in healthcare," Inf. Fusion, vol. 114, Art. no. 102690, 2025, doi: 10.1016/j.inffus.2024.102690.
12. B. Li et al., "Trustworthy AI: From principles to practices," ACM Comput. Surv., vol. 55, no. 9, pp. 1–46, 2023, doi: 10.1145/3555803.

13. S. Ali et al., "Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence," Inf. Fusion, vol. 99, Art. no. 101805, 2023, doi: 10.1016/j.inffus.2023.101805.
14. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst., Man, Cybern., vol. SMC-3, no. 6, pp. 610–621, 1973, doi: 10.1109/TSMC.1973.4309314.
15. T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, 2002, doi: 10.1109/TPAMI.2002.1017623.
16. F. Makhmudov, J. Khamzaev, M. Hudayberdiev, B. Achilov, S. Otamuradov, T. Kuchkorov, I. Saymanov, and A. Kutlimuratov, "Fractal-based image analysis for multi-stage detection of tomato late blight using a laboratory image dataset of greenhouse-grown tomato plants," Horticulturae, vol. 12, Art. no. 979, 2026, doi: 10.3390/horticulturae12080979.
17. A. Kutlimuratov, B. Achilov, K. Seitnazarov, P. Allayarov, I. Saymanov, R. Oteniyazov, and J. Khamzaev, "XGBoost ensemble algorithm for classifying tomato leaf diseases based on texture descriptors," AgriEngineering, vol. 8, Art. no. 98, 2026, doi: 10.3390/agriengineering8030098.
18. F. Safarov, A. Kutlimuratov, U. Khojamuratova, A. Abdusalomov, and Y.-I. Cho, "Enhanced AlexNet with Gabor and local binary pattern features for improved facial emotion recognition," Sensors, vol. 25, Art. no. 3832, 2025, doi: 10.3390/s25123832.
19. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
20. F. Makhmudov, S. Umirzakova, A. Kutlimuratov, A. Abdusalomov, and Y.-I. Cho, "Advanced object detection for maritime fire safety," Fire, vol. 7, Art. no. 430, 2024, doi: 10.3390/fire7120430.
21. A. Mamadmurodov, S. Umirzakova, M. Rakhimov, A. Kutlimuratov, Z. Temirov, R. Nasimov, A. Meliboev, A. Abdusalomov, and Y. Im Cho, "A hybrid deep learning model for early forest fire detection," Forests, vol. 16, Art. no. 863, 2025, doi: 10.3390/f16050863.
22. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), in Proc. Mach. Learn. Res., vol. 97, 2019, pp. 6105–6114.
23. A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.
24. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol. (NAACL-HLT), vol. 1, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
25. Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2021, pp. 9992–10002, doi: 10.1109/ICCV48922.2021.00986.
26. J. Chen et al., "TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," Med. Image Anal., vol. 97, Art. no. 103280, 2024, doi: 10.1016/j.media.2024.103280.
27. A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," in Brainlesion: Glioma,

Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes 2021), in Lecture Notes in Computer Science, vol. 12962, 2022, pp. 272–284, doi: 10.1007/978-3-031-08999-2_22.
28. F. Shamshad et al., "Transformers in medical imaging: A survey," Med. Image Anal., vol. 88, Art. no. 102802, 2023, doi: 10.1016/j.media.2023.102802.
29. J. Khamzaev, J. Karimberdiyev, M. Rakhimov, I. Saymanov, S. Otamurodov, O. Rikhsimboev, I. Dmitriy, A. Kutlimuratov, and F. Makhmudov, "Hybrid Transformer–CNN with boundary-aware attention for accurate multi-modal brain tumor segmentation," BioMedInformatics, vol. 6, Art. no. 46, 2026, doi: 10.3390/biomedinformatics6040046.
30. A. Kutlimuratov, K. Sharipov, P. Allayarov, S. Iskandarova, R. Latyfskiy, G. Tolibaeva, and F. Makhmudov, "SE-POSTER: Channel-enhanced landmark guided Transformer for facial emotion recognition," Informatics, vol. 13, Art. no. 123, 2026, doi: 10.3390/informatics13080123.
31. T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learn. Represent. (ICLR), 2017. [Online]. Available: arXiv:1609.02907
32. A. Kutlimuratov, I. Saymanov, B. Khujamuratov, F. Makhmudov, T. Djurayev, N. Irmukhamedova, H. Adinaev, F. Sarikulova, and K. Khudaybergenov, "Action-unit-constrained graph learning with counterfactual recourse for auditable facial expression recognition," Computers, vol. 15, Art. no. 750, 2026, doi: 10.3390/computers15110750.
33. G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv:1503.02531, 2015.
34. K. Sun, X. Wang, X. Miao, and Q. Zhao, "A review of AI edge devices and lightweight CNN and LLM deployment," Neurocomputing, vol. 614, Art. no. 128791, 2025, doi: 10.1016/j.neucom.2024.128791.
35. A. Abdusalomov, S. Umirzakova, A. Kutlimuratov, D. Mirzaev, A. Dauletov, T. Botirov, M. Zakirova, M. Mukhiddinov, and Y. I. Cho, "Lightweight UAV-based system for early fire-risk identification in wild forests," Fire, vol. 8, Art. no. 288, 2025, doi: 10.3390/fire8080288.
36. A. Kutlimuratov, F. Rakhmatov, J. Khamzaev, I. Saymanov, P. Allayarov, G. Bekbaev, S. Otamurodov, and F. Makhmudov, "A lightweight cascade-based framework for real-time zero-day attack detection," Computers, vol. 15, Art. no. 174, 2026, doi: 10.3390/computers15030174.
37. A. Kutlimuratov, I. Saymanov, B. Khujamuratov, D. Uskenbaeva, F. Makhmudov, M. Rakhmanov, U. Bekmurodov, J. Karimboev, and K. Khudaybergenov, "Compressed-domain temporal cues and degradation-aware distillation for lightweight dynamic facial expression recognition on edge devices," Technologies, vol. 14, Art. no. 729, 2026, doi: 10.3390/technologies14110729.
38. A. V. Geetha, T. Mala, D. Priyanka, and E. Uma, "Multimodal emotion recognition with deep learning: Advancements, challenges, and future directions," Inf. Fusion, vol. 105, Art. no. 102218, 2024, doi: 10.1016/j.inffus.2023.102218.
39. A. Abdusalomov, M. Mukhiddinov, K. Abdurashidova, A. Kutlimuratov, A. Marakhimov et al., "SYMPHONIA—Enhanced multimodal emotion recognition with dual-branch dynamic attention and hierarchical adaptive fusion," Comput. Mater. Contin., vol. 88, no. 1, Art. no. 74, 2026, doi: 10.32604/cmc.2026.077057.
40. A. Abdusalomov, M. Mukhiddinov, F. Abdirazakov, A. Kutlimuratov, N. Alimova et al., "A prosody-guided multi-stream framework for universal detection of AI-synthesized speech across codec and

vocoder domains," Comput. Mater. Contin., vol. 88, no. 1, Art. no. 98, 2026, doi: 10.32604/cmc.2026.080444.
41. Y.-W. Chao, Y. Liu, X. Liu, H. Zeng, and J. Deng, "Learning to detect human-object interactions," in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), 2018, pp. 381–389, doi: 10.1109/WACV.2018.00048.
42. A. Abdusalomov, U. Xudoynazarov, U. Ruziev et al., "Multi-view context learning for human-object interaction detection," Sci. Rep., vol. 16, Art. no. 32684, 2026, doi: 10.1038/s41598-026-71464-8.
43. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 27, 2014, pp. 2672–2680.
44. J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020, pp. 6840–6851.
45. A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep learning approaches for data augmentation in medical imaging: A review," J. Imaging, vol. 9, no. 4, Art. no. 81, 2023, doi: 10.3390/jimaging9040081.
46. M. J. A. Rasool, A. Abdusalomov, A. Kutlimuratov, M. J. A. Ahamed, S. Mirzakhalilov, A. Shavkatovich Buriboev, and H. S. Jeon, "PixMed-Enhancer: An efficient approach for medical image augmentation," Bioengineering, vol. 12, Art. no. 235, 2025, doi: 10.3390/bioengineering12030235.
47. K. Singhal et al., "Large language models encode clinical knowledge," Nature, vol. 620, no. 7972, pp. 172–180, 2023, doi: 10.1038/s41586-023-06291-2.
48. P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020, pp. 9459–9474.
49. C. Zakka et al., "Almanac — Retrieval-augmented language models for clinical medicine," NEJM AI, vol. 1, no. 2, Art. no. AIoa2300068, 2024, doi: 10.1056/AIoa2300068.
50. M. Li, Y. Ahmadiadli, and X.-P. Zhang, "A survey on speech deepfake detection," ACM Comput. Surv., vol. 57, no. 7, pp. 1–38, 2025, doi: 10.1145/3714458.
51. B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," IEEE Trans. Med. Imag., vol. 34, no. 10, pp. 1993–2024, 2015, doi: 10.1109/TMI.2014.2377694.
52. P. Zhu et al., "Detection and tracking meet drones challenge," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 11, pp. 7380–7399, 2022, doi: 10.1109/TPAMI.2021.3119563.
53. X. Wang et al., "ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech," Comput. Speech Lang., vol. 64, Art. no. 101114, 2020, doi: 10.1016/j.csl.2020.101114.
54. A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, Art. no. 20, 2019, doi: 10.1186/s42400-019-0038-7.
55. A. Abdusalomov, K. Zohirov, A. Khojamurotov, F. Safarov, A. Kutlimuratov et al., "An NLP-based neuro-semantic clinical filter for medical text simplification," Comput. Mater. Contin., vol. 88, no. 3, Art. no. 25, 2026, doi: 10.32604/cmc.2026.079237.
56. J. de Lope and M. Graña, "An ongoing review of speech emotion recognition," Neurocomputing, vol. 528, pp. 1–11, 2023, doi: 10.1016/j.neucom.2023.01.002.

57. I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," in Neural Information Processing (ICONIP 2013), in Lecture Notes in Computer Science, vol. 8228, 2013, pp. 117–124, doi: 10.1007/978-3-642-42051-1_16.
58. P. V. A. B. de Venâncio, A. C. Lisboa, and A. V. Barbosa, "An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices," Neural Comput. Appl., vol. 34, no. 18, pp. 15349–15368, 2022, doi: 10.1007/s00521-022-07467-z.
59. I. Pacal et al., "A systematic review of deep learning techniques for plant diseases," Artif. Intell. Rev., vol. 57, no. 11, Art. no. 304, 2024, doi: 10.1007/s10462-024-10944-7.
60. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in Proc. IEEE Symp. Comput. Intell. Secur. Defense Appl. (CISDA), 2009, pp. 1–6, doi: 10.1109/CISDA.2009.5356528.
61. A. Abdusalomov, S. Umirzakova, O. Bekmirzaev, A. Dauletov, A. Buriboev, A. Kutlimuratov, A. Nishanov, R. Nasimov, and R. Oh, "From anatomy to genomics using a multi-task deep learning approach for comprehensive glioma profiling," Bioengineering, vol. 12, Art. no. 979, 2025, doi: 10.3390/bioengineering12090979.
62. A. Kutlimuratov, D. Eshmurodov, F. Tulaganova, A. Utegenov, P. Allayarov, J. Khamzaev, I. Saymanov, and F. Makhmudov, "Early diagnosis of blood disorders via enhanced image preprocessing and deep learning modeling," BioMedInformatics, vol. 6, Art. no. 25, 2026, doi: 10.3390/biomedinformatics6030025.
63. S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," Patterns, vol. 4, no. 9, Art. no. 100804, 2023, doi: 10.1016/j.patter.2023.100804.
64. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
65. A. Abdusalomov, S. Umirzakova, S. Mirzakhalilov, A. Kutlimuratov, R. Nasimov, Z. Temirov, W. Jeong, H. Choi, and T. K. Whangbo, "A generative expert-narrated simplification model for enhancing health literacy among the older population," Bioengineering, vol. 12, Art. no. 1066, 2025, doi: 10.3390/bioengineering12101066.
66. R. Nasimov, A. Abdusalomov, C. Khidirova, K. Temirova, A. Kutlimuratov, S. Sadikova, W. Jeong, H. Choi, and T. K. Whangbo, "PatientEase—Domain-aware RAG for rehabilitation instruction simplification," Bioengineering, vol. 12, Art. no. 1204, 2025, doi: 10.3390/bioengineering12111204.
67. C. Busso et al., "IEMOCAP: Interactive emotional dyadic motion capture database," Lang. Resour. Eval., vol. 42, no. 4, pp. 335–359, 2008, doi: 10.1007/s10579-008-9076-6.
68. A. Abdusalomov, A. Kutlimuratov, R. Nasimov, and T. K. Whangbo, "Improved speech emotion recognition focusing on high-level data representations and swift feature extraction calculation," Comput. Mater. Contin., vol. 77, no. 3, pp. 2915–2933, 2023, doi: 10.32604/cmc.2023.044466.
69. S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual explanations without opening the black box: Automated decisions and the GDPR," Harv. J. Law Technol., vol. 31, no. 2, pp. 841–887, 2018.
70. M. U. Haq, M. A. J. Sethi, S. Ahmad, N. Ahmad, M. S. Anwar et al., "A comprehensive review of face detection/recognition algorithms and competitive datasets to optimize machine vision," Comput. Mater. Contin., vol. 84, no. 1, pp. 1–24, 2025, doi: 10.32604/cmc.2025.063341.

71. S. P. H. Boroujeni et al., "A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management," Inf. Fusion, vol. 108, Art. no. 102369, 2024, doi: 10.1016/j.inffus.2024.102369.
72. A. Saleh, M. A. Zulkifley, H. H. Harun, F. Gaudreault, I. Davison, and M. Spraggon, "Forest fire surveillance systems: A review of deep learning methods," Heliyon, vol. 10, no. 1, Art. no. e23127, 2024, doi: 10.1016/j.heliyon.2023.e23127.
73. I. Shamta and B. E. Demir, "Development of a deep learning-based surveillance system for forest fire detection and monitoring using UAV," PLoS ONE, vol. 19, no. 3, Art. no. e0299058, 2024, doi: 10.1371/journal.pone.0299058.
74. B. Khujamuratov, K. Abdurashidova, A. Kutlimuratov, and T. Khudaybergenov, "A leakage-aware evaluation framework and benchmark for edge-based multi-sensor fire and smoke detection," Information, vol. 17, Art. no. 1032, 2026, doi: 10.3390/info17101032.
75. J. Wang, W. Yang, H. Guo, R. Zhang, and G.-S. Xia, "Tiny object detection in aerial images," in Proc. 25th Int. Conf. Pattern Recognit. (ICPR), 2021, pp. 3791–3798, doi: 10.1109/ICPR48806.2021.9413340.
76. R. Nasimov, J. Sevinov, A. Abdusalomov, Y. Oğurlu, N. Nosir, A. Abdullaev, A. Kutlimuratov, K. Avazov, and Y. I. Cho, "Measure-based supervision for tiny object detection in aerial imagery," Computers, vol. 15, Art. no. 744, 2026, doi: 10.3390/computers15110744.
77. K. Avazov, K. Zohirov, A. Kutlimuratov, C. Khidirova, J. Sevinov et al., "Context-aware identity validation for UAV-assisted urban mobility and traffic monitoring environments," Comput. Mater. Contin., vol. 88, no. 3, Art. no. 69, 2026, doi: 10.32604/cmc.2026.083828.
78. D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," arXiv:1511.08060, 2015.
79. T. Nyawose, R. C. Maswanganyi, and P. Khumalo, "A review on the detection of plant disease using machine learning and deep learning approaches," J. Imaging, vol. 11, no. 10, Art. no. 326, 2025, doi: 10.3390/jimaging11100326.
80. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Trans. Evol. Comput., vol. 6, no. 2, pp. 182–197, 2002, doi: 10.1109/4235.996017.
81. Z. Kalpeyeva, Z. Abildaeva, R. Uskenbayeva, A. Kassymova, A. Kutlimuratov, P. Allayarov, A. Abdusalomov, and Y.-I. Cho, "A hybrid NSGA-II and machine learning framework for multi-objective marketing budget optimization in Kazakhstan's agro-industrial complex," Sustainability, vol. 18, Art. no. 7632, 2026, doi: 10.3390/su18157632.
82. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy (ICISSP), 2018, pp. 108–116, doi: 10.5220/0006639801080116.
83. M. M. I. Munna, M. M. Rahman, J. Frnda et al., "Elevating intrusion detection and security fortification in intelligent networks through cutting-edge machine learning paradigms," Sci. Rep., vol. 15, Art. no. 39989, 2025, doi: 10.1038/s41598-025-23754-w.
84. Y. Guo, "A review of machine learning-based zero-day attack detection: Challenges and future directions," Comput. Commun., vol. 198, pp. 175–185, 2023, doi: 10.1016/j.comcom.2022.11.001.

85. S. Gulomov, S. Jumayev, S. Bozorov, I. Boykuziev, A. Kutlimuratov, and I. Saymanov, "Integrated anomaly detection and mitigation in SDN environments: A hybrid approach," Computers, vol. 15, Art. no. 641, 2026, doi: 10.3390/computers15090641.
86. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
87. B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," in Proc. Int. Conf. Learn. Represent. (ICLR), 2016. [Online]. Available: arXiv:1511.06939
88. P. Covington, J. Adams, and E. Sargin, "Deep neural networks for YouTube recommendations," in Proc. 10th ACM Conf. Recommender Syst. (RecSys), 2016, pp. 191–198, doi: 10.1145/2959100.2959190.
89. A. Ilyosov, A. Kutlimuratov, and T.-K. Whangbo, "Deep-sequence–aware candidate generation for e-learning system," Processes, vol. 9, Art. no. 1454, 2021, doi: 10.3390/pr9081454.
90. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 4765–4774.