

PROJECT REPORT ON

Cardiology Diagnosis Automation

Zhurek AI

Submitted by:

Ariana Kenbayeva
Kymbat Amanova
Amina Aimuratova
Ablay Abdimalinov
Zhazelya Muratkhanova

Supervised by:

Dr. Siamac Fazli
Department of Computer Science, SEDS

CSCI-409 Senior Project

Department of Computer Science
School of Engineering and Digital Sciences
Nazarbayev University, Astana, Kazakhstan

Spring 2026

Contents

1	Executive Summary	1
1.1	Problem Statement and Objectives	1
1.2	Design, Implementation, and Evaluation Summary	1
2	Introduction	2
2.1	Significance of the Project	2
2.2	Brief Overview of the Solution	2
3	Identification of Constraints	3
3.1	Constraints Affecting the Project	3
4	Background and Related Work	3
4.1	Literature Review	4
4.1.1	EchoCV	4
4.1.2	EchoCLIP	4
4.1.3	PanEcho	4
4.1.4	EchoPrime	4
4.1.5	Commercial Systems	5
4.2	Comparative Analysis	5
5	Project Approach	7
5.1	System Overview	7
5.2	Software Architecture	8
5.2.1	Node.js + Express Backend	8
5.2.2	Database Schema	9
5.2.3	React + TypeScript Frontend	11
5.3	ML Pipeline and Algorithms	12
5.3.1	EchoPrime Architecture	12
5.3.2	GradCAM Interpretability	12
5.3.3	Algorithmic Evolution	12
5.4	User Workflow and Features	12
5.4.1	Use Cases	12
5.4.2	Features Summary	14
5.5	Technology Stack	15
6	Project Execution	15
6.1	Project Timeline Across Two Semesters	15
6.2	Major Design Decisions	16
6.3	Changes in Project Scope or Direction	16
6.4	Challenges Encountered	16
6.5	Solutions and Problem Handling	17
6.6	Teamwork and Collaboration	17

7	Evaluation	17
7.1	Evaluation Objectives	17
7.2	Evaluation Methodology	18
7.3	Validation of the Proposed Solution	18
7.4	Discussion of Findings	18
8	Conclusion and Future Work	19
8.1	Conclusion	19
8.2	Future Work	19
9	Ethical and Legal Considerations	20
9.1	Identification of Ethical Issues	20
9.2	Legal Compliance	20
9.3	Sustainability and Social Responsibility	20
A	Glossary	23
	Appendix A: Glossary	23

1. Executive Summary

1.1 Problem Statement and Objectives

The leading cause of deaths in Kazakhstan is Cardiovascular disease(CDV). Some revisions were made in international heart-failure guidelines since 2016, but new diagnostic standards are adopted slowly in primary care, especially outside main cities [1]. This gap is caused by a noticeable shortage of specialists capable of interpreting echocardiographic (echo) studies.

Based on this, the team identified three key objectives:

1. Lessen the reliance on specialists by automating echocardiographic analysis.
2. Lower the amount of misinterpretation of important cardiac measurements through AI-driven report generation.
3. Ensure clinical transparency through explainable and verifiable AI outputs that clinicians can trust.

To address these objectives the team built **Zhurek AI**, an AI-powered web platform for echocardiography analysis. The platform creates structured clinical reports covering [ejection fraction](#), valvular evaluation, wall motion, and pericardial findings. It is powered by **EchoPrime** [2], a multi-video view-informed vision-language model trained on 12.1 million echocardiographic clips, which achieves a mean AUC of 0.85–0.92 and exceeds human cardiologists on several tasks.

To support clinician trust, the platform provides:

- Per-frame [GradCAM](#) saliency heatmaps overlaid on [DICOM](#) footage.
- **Editable report fields** so physicians can correct AI-generated text before sign-off.
- A **physician sign-off workflow** enforced at both application and database level.

These features were validated by a structured usability session with a practising cardiac sonographer, whose feedback directly resulted in three specific feature additions.

1.2 Design, Implementation, and Evaluation Summary

Zhurek AI is a full-stack computing solution spanning all three stages of the software engineering lifecycle.

Design. Four candidate model architectures were evaluated (EchoCV, EchoCLIP, PanEcho, EchoPrime). EchoPrime was selected for integration. Domain-driven design with an aggregate-root pattern was adopted to enforce clinical state machines at both the

application and database levels.

Implementation. The system was realised as three independentl deployable services: a React + TypeScript frontend; a FastAPI backend deployed on Render; and a Python ML inference backend wrapping the EchoPrime API. DICOM files are transmitted straight from the browser to the ML backend, bypassing the main application server entirely for privacy reasons. A custom GradCAM module was implemented on top of EchoPrime to restore spatial interpretability.

Evaluation. The integrated platform was assessed by scheduling a usability session with a practising cardiac sonographer. Feedback from that drove three feature additions: GradCAM visualisations, editable report fields, and Word-document export.

2. Introduction

2.1 Significance of the Project

Cardiovascular disease is the most frequent cause of death in Kazakhstan. Between 2013 and 2020, diagnosed circulatory-system cases grew from 18 683 to 26 306 per 100 000 people [3]. Deaths caused by heart failure almost tripled between 2014 and 2019, summing to about 2.3 million disability-adjusted life years in 2019 alone [4].

The main driver of late diagnosis is that there are not enough specialists. In rural areas, primary-care physicians are rarely trained to interpret cardiac ultrasound. Even in urban hospitals, a cardiologist may review dozens of studies per day; a standard echo examination takes 26 minutes to perform and up to 20 minutes to report on [6, 7], taking roughly an hour of specialist time per patient.

The diagnosis that is most often missed in this context is **heart failure with preserved ejection fraction (HFpEF)**: the heart contracts normally but cannot relax and fill between beats. Because the ejection fraction looks normal, an undertrained clinician could discharge the patient without a diagnosis. Detecting HFpEF needs close scrutiny of multiple echo parameters aside from ejection fraction — a skill uncommon outside major centres. This is exactly the diagnostic gap that automated echo interpretation can close.

2.2 Brief Overview of the Solution

Zhurek AI is an AI-driven web application designed to close this diagnostic gap. It does not replace cardiologists; it provides available physicians — including those in district hospitals — with a systematic, comprehensive echo report that is generated automatically.

The platform uses **EchoPrime**, a multimodal vision-language model to:

- Automatically label cardiac views in uploaded DICOM files.
- Produce measurements: ejection fraction, valvular assessment, wall motion, and pericardial findings.

A **GradCAM** extension generates per-frame saliency heatmaps, allowing clinicians to verify which image regions drove each AI assessment before signing off the report.

3. Identification of Constraints

3.1 Constraints Affecting the Project

Economic. Development depended entirely on free-tier cloud services, restricting GPU access, storage, and bandwidth. There was no budget allocated for local data collection or large-scale model training. Wide clinical deployment would require funding from hospitals, universities, or government.

Social. The system depends on clinician trust. There is also a risk that deploying only in well-funded urban hospitals could enlarge existing gap between urban and rural regions.

Political and Regulatory. Clinical AI systems must satisfy approval requirements from Kazakhstan’s Ministry of Health. Because the backend is hosted in the EU, cross-border data transfers must comply with GDPR and Kazakhstan’s data-protection regulations.

Ethical. Patient data must be deidentified prior to upload. The system is designed to assist, not replace clinical judgment. Model training must not use uploaded patient data without explicit consent.

Health and Safety. Incorrect AI outputs can delay diagnosis, cause unnecessary follow-up tests or incorrect treatment decisions, so all AI-generated reports are treated as draft outputs requiring physician review and sign-off before clinical use.

4. Background and Related Work

4.1 Literature Review

CVD is the main cause of mortality globally, and the number is expected to grow — from approximately 20.5 million deaths in 2025 to 35.6 million by 2050, with ischaemic heart disease the leading contributor [8].

Echocardiography is the most widely used cardiac imaging tool. It carries no radiation exposure, can be performed at the bedside, and provides real-time structural and functional data [9, 10]. A study consists of multiple video clips captured from different chest positions, each providing a different cardiac view; a written report synthesises these views into a clinical summary.

Recent advances in medical AI have produced vision-language models that learn joint representations of echo videos and clinical text. **Contrastive Language-Image Pre-training (CLIP)** [11] trains two encoders — one for images, one for text — so that matched image-text pairs are drawn together in a shared embedding space. Applied to echocardiography, this means an echo clip can be semantically aligned with the cardiologist report that describes it.

4.1.1 *EchoCV*

EchoCV [12] introduced automated view classification using a VGG13 network to categorise echocardiographic frames into 23 views. It provides a useful benchmark but outputs only a view label — no measurements, no report, and no diagnostic coding. It also lacks the Doppler and zoomed sub-views common in large datasets such as MIMIC-IV-ECHO.

4.1.2 *EchoCLIP*

EchoCLIP [13] applied CLIP-style vision-language learning to echocardiography, enabling zero-shot inference on tasks not explicitly trained on. However, its generation is retrieval-based rather than generative: it finds the closest existing report in a corpus rather than creating a new, patient-specific narrative. It also operates on single clips rather than full multi-view studies.

4.1.3 *PanEcho*

PanEcho [14] predicts 39 clinical outputs simultaneously (both classifications and continuous measurements including **LVEF** and chamber dimensions) and was validated on large cohorts. Its key weakness is a *view-agnostic* design: applying the same inference to any DICOM clip regardless of anatomical window produces intra-study measurement variance of 10–30%, which is clinically unacceptable.

4.1.4 *EchoPrime*

EchoPrime [2] is the current state of the art in academic echocardiography AI. It accepts a full multi-view study as input and applies a view-informed anatomical attention mechanism: it first classifies each clip’s view, then weights each clip’s contribution to interpretation of a given cardiac structure. This allows to mimic real clinician reasoning. Attention weights

are anatomically interpretable. EchoPrime achieved a mean AUC of 0.85–0.92 through five hospitals in three countries and outperformed previous foundation models and in some comparisons human cardiologists.

4.1.5 Commercial Systems

Us2.ai [15] used for view recognition, measurement extraction, disease detection and report generation in a standard workflow with FDA approval. It shows that it works in real clinical settings. However, it is closed-source and does not include automatic ICD-10 coding.

Ultromics EchoGo [16, 17] gives a wide range of measurements and disease-specific features but it also does not have ICD-10 coding and relies on cloud-only architecture, which could have data-sovereignty concerns.

4.2 Comparative Analysis

Table 1 sums up s the key properties of mentioned systems.

Table 1. Comparative analysis of existing echo-to-report systems.

nublu							
EchoCV (2018)	DICOM frames	View label (23 classes)	Yes	No	No	No	measurements or report; low Doppler sub-view accuracy
EchoCLIP (2024)	Video+text pairs	Retrieved report embeddings	No	Partial	No	Retrieval only; no new text; single-clip only	
PanEcho (2025)	Any DICOM clip	39 measurements	No	No	No	10–30% intra-study variance; no narrative; no ICD	
EchoPrime (2026)	Full multi-view study	Retrieved+synthesized report	Yes	Partial	No	Retrieval-based; large reference corpus; no ICD	
Us2.ai	Any DICOM	Structured report + disease flags	Yes	Partial	No	Proprietary; no ICD allocation	
Ultramics EchoGo	Cloud DICOM	Structured report + disease flags	Yes	Partial	No	Proprietary; cloud-only; no ICD	
Zhurek AI (ours)	DICOM study	Structured report + GradCAM	Yes	Yes	Planned	No local KZ fine-tuning; ICD pending	

5. Project Approach

5.1 System Overview

Zhurek AI (zhurek-frontend.onrender.com) is built with three services communicating over HTTP:

1. **React + TypeScript Frontend** (Vite): browser-based clinical interface.
2. **Node.js + Express Backend** (Render, PostgreSQL): authentication, patient records, studies, report lifecycle, processes metadata only.
3. **Python FastAPI ML Backend** (Hugging Face Space): EchoPrime inference and GradCAM serving.

Key architectural decision: DICOM files never reach the Node.js backend. They are loaded into browser memory by `file-store.ts` and transmitted directly to the Hugging Face Space. The Node.js backend takes only study metadata (date, external patient ID, study UID) and fires an analysis trigger carrying those identifiers, not the binary data. When analysis completes, the ML backend calls a `PATCH` endpoint on the Node.js side to persist results. This keeps the backend stateless with respect to binary data and the need for persistent disk storage on the free-tier Render environment gets removed.

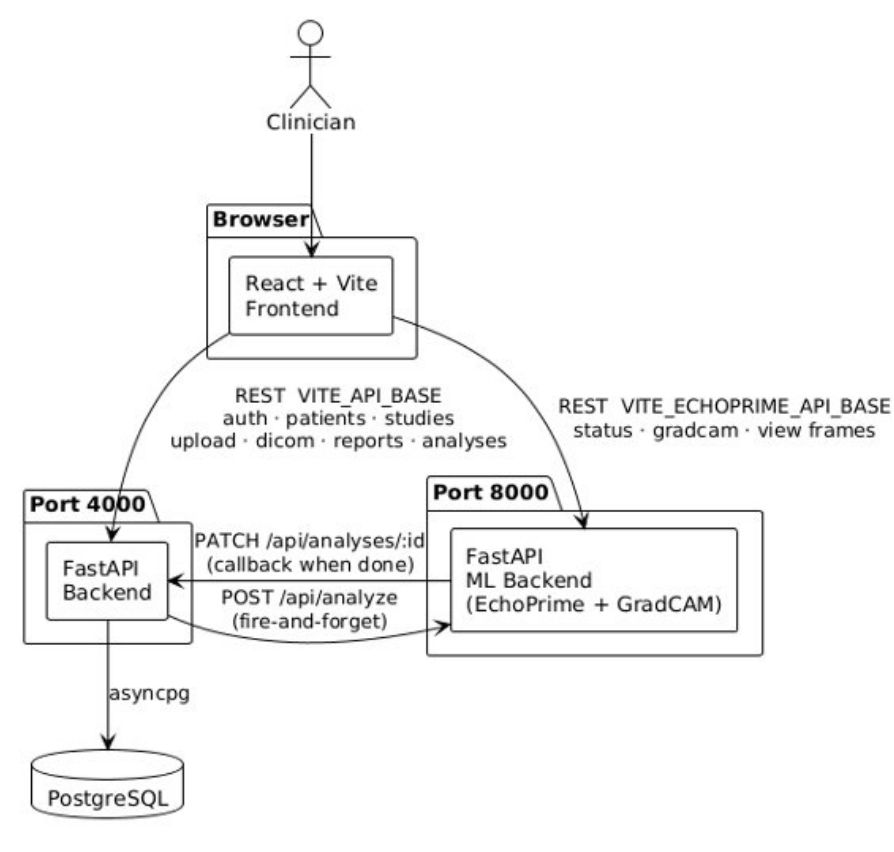


Figure 1. System architecture: three-service deployment with direct browser-to-ML DICOM routing.

5.2 Software Architecture

5.2.1 Node.js + Express Backend

The backend is organised in these two layers:

- **Route handlers** are thin adapters: deserialise the HTTP request, invoke one domain function, return the result.
- **Domain modules** are designed as Abstract Data Types (ADTs), each encapsulating a certain responsibility:
 - *Identity*: JWT encoding, bcrypt hashing, token validation
 - *Patient*: pseudonym generation and deduplication
 - *DICOMCorpus*: DICOM filename metadata
 - *EchoPipeline*: HTTP call and retry logic to the ML service
 - *AnalysisResult*: pure value type over inference outputs
 - *ClinicalReport*: draft/finalized status-machine enforcement
 - *EchoStudy*: aggregate root coordinating five tables, access control, status

transitions, and corpus/pipeline/report interaction

Module contracts are enforced with Design by Contract (DbC): preconditions state caller guarantees; postconditions state callee promises. A global exception handler maps contract violations to HTTP 400, permission errors to 403, and not-found errors to 404.

5.2.2 Database Schema

The seven tables of PostgreSQL schema are: `users`, `user_consent`s, `patients`, `consent`s, `studies`, `study_uploads`, `analyses`, and `reports`. All primary keys are UUIDs.

Key integrity constraints:

- `studies.status` $\in$ {`queued`, `analysing`, `complete`, `failed`}
- `reports.status` $\in$ {`draft`, `finalized`}
- `analyses` requires `result` IS NOT NULL when `status` = `'done'`
- `reports` requires `finalized_at` IS NOT NULL when `status` = `'finalized'`

Report finalization is **irreversible by design**: a PostgreSQL trigger (`enforce_report_finalization`) blocks any `UPDATE` that attempts to move status away from `finalized`, giving a second enforcement layer separate from the application.

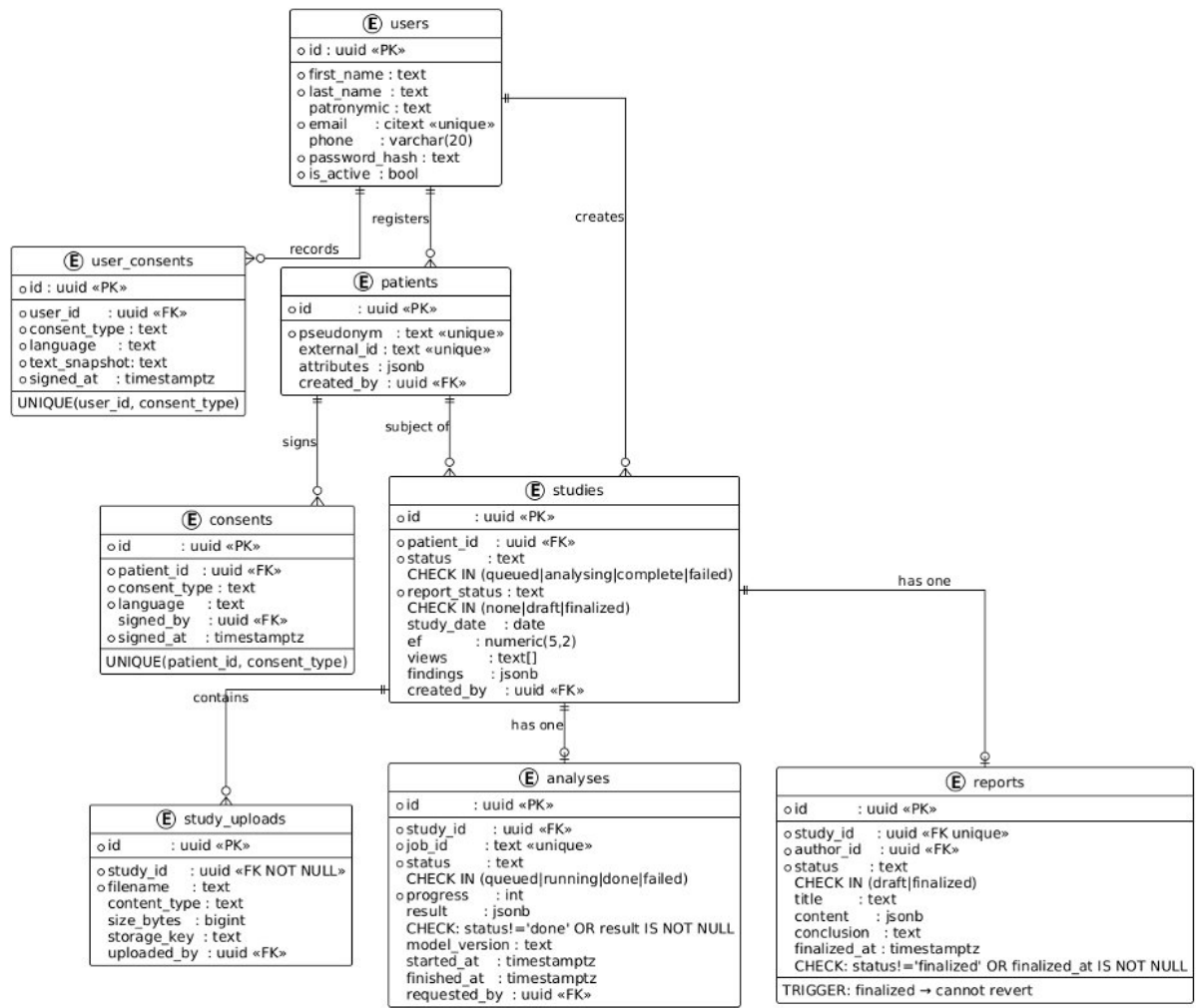


Figure 2. PostgreSQL database schema showing seven tables and their relationships.

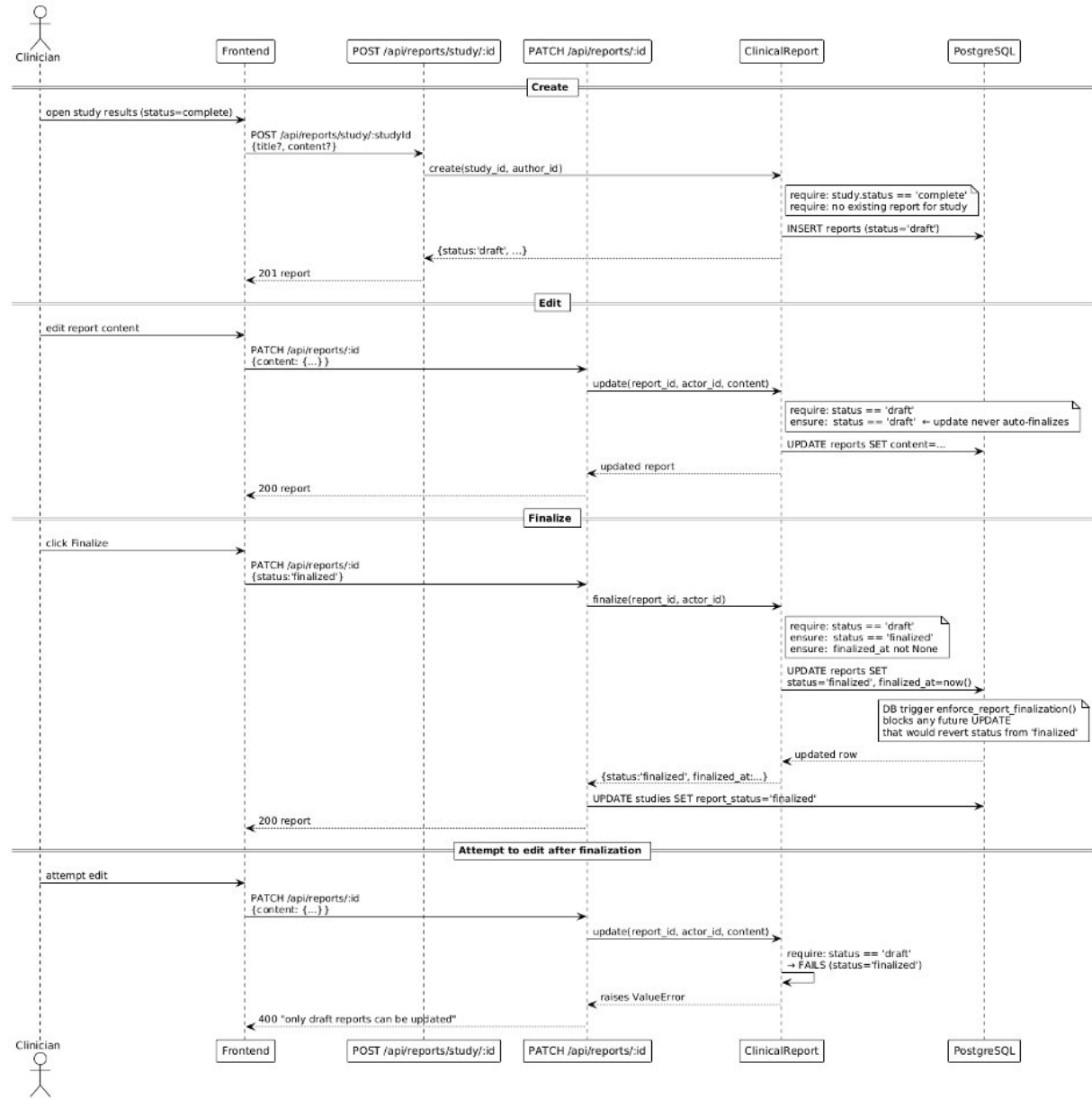


Figure 3. Report lifecycle sequence: create, edit, finalize, and attempted post-finalization edit (blocked).

5.2.3 React + TypeScript Frontend

The frontend is a single-page application (SPA). It was built using React, TypeScript, and Vite. Protected routes made is so that a clinician must complete an ethics consent form before accessing and editing the studies.

The DICOM viewer is built on Cornerstone.js and supports playback speed control, frame inversion, and GradCAM overlay mode. When the overlay is active, heatmap frames are pre-decoded to `ImageBitmap` objects and drawn synchronously to a `<canvas>` on each frame tick, bypassing React's render cycle to maintain frame synchrony at high playback rates.

5.3 ML Pipeline and Algorithms

5.3.1 *EchoPrime Architecture*

The ML backend runs EchoPrime, a pre-existing open-source multi-task cardiac AI model. Four components operate in sequence:

1. **MViT v2 encoder** produces 512-dimensional video embeddings per DICOM clip.
2. **ConvNeXt view classifier** labels each clip by echocardiographic view (e.g. apical four-chamber, parasternal long axis).
3. **Report generation transformer** produces structured text from embeddings.
4. **MLP metric predictor** outputs quantitative measures including ejection fraction.

EchoPrime’s training objective is multimodal: echo videos are encoded via MViT v2 while cardiology reports are encoded via BioMedBERT, and both modalities are projected into a shared latent space using a CLIP-style contrastive objective.

5.3.2 *GradCAM Interpretability*

The video encoder compresses spatial information into dense embeddings, so clinicians cannot directly observe why the model reached a particular conclusion. To address this, the team implemented a custom GradCAM attribution module: gradient hooks on the MViT v2 encoder produce per-frame saliency heatmaps, composited as RGBA overlays over the original DICOM frames and returned to the frontend as base64-encoded JPEG sequences.

5.3.3 *Algorithmic Evolution*

Early exploration included a multi-stage pipeline with separate view classification, segmentation, and disease-specific prediction modules, as well as a PanEcho-style view-agnostic direction. The team consolidated around EchoPrime because it better suited the goal of producing richer, report-style interpretation from echo videos rather than isolated classification outputs.

5.4 User Workflow and Features

5.4.1 *Use Cases*

Two main actors in the platform:

- **Clinician** (cardiologist or sonographer)
- **System** (automated ML pipeline)

Primary use cases:

1. **Register and consent:** clinician creates an account and completes the mandatory ethics consent form.

2. **Register patient:** clinician registers a patient using a pseudonymised identifier.
3. **Upload study:** DICOM files are routed straight to the ML service; metadata is registered with the Node.js backend.
4. **Analyse study:** system classifies views, estimates cardiac metrics, and generates a structured report.
5. **Review results:** clinician inspects the structured report, cardiac metrics, and DICOM viewer with optional GradCAM overlay.
6. **Create and edit report:** clinician creates a draft, edits generated text, and may validate model attention via GradCAM heatmaps.
7. **Finalise and export report:** clinician irreversibly finalises the report and exports it as a Word document.

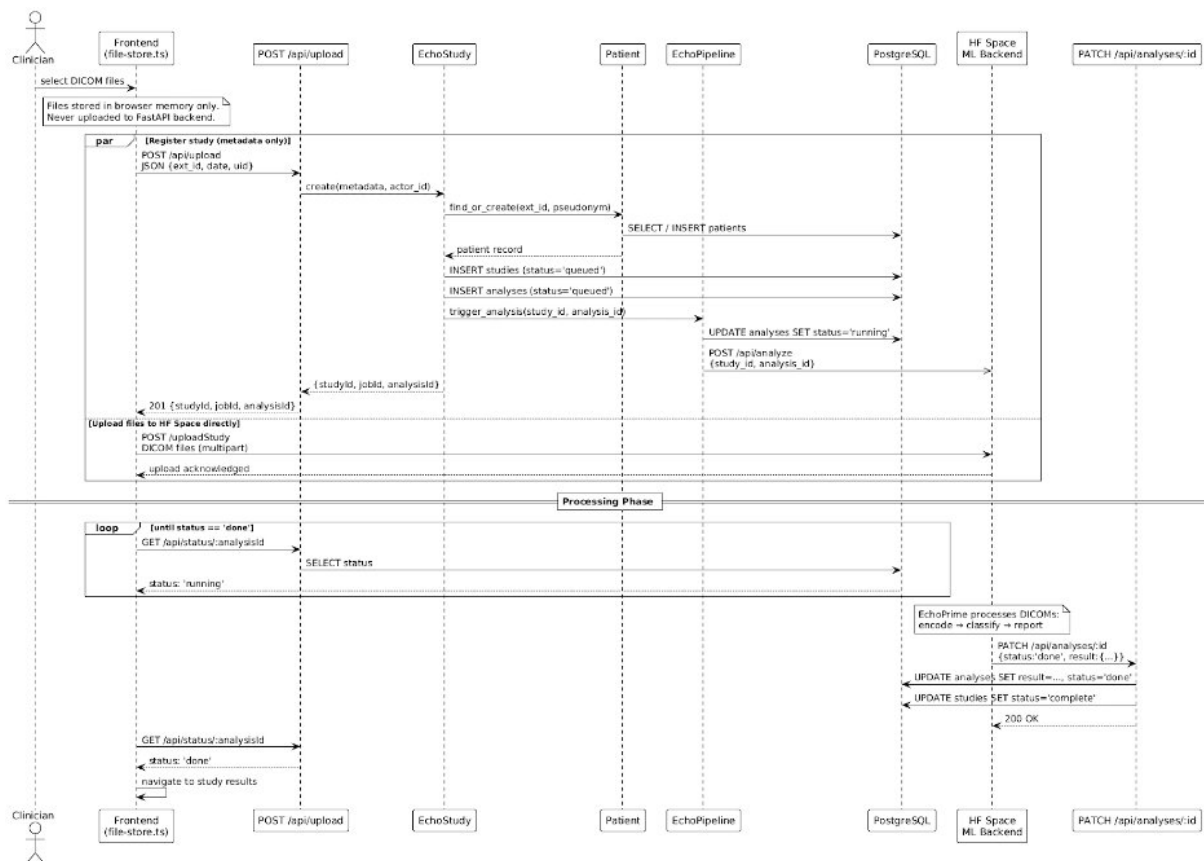


Figure 4. Clinical workflow sequence: DICOM upload, analysis trigger, processing phase, and results delivery.

5.4.2 Features Summary

Table 2. Platform feature summary.

nublu	
Study upload	Browser-based DICOM ingestion with direct ML service routing; main backend never receives files
Automated analysis	View classification, ejection fraction estimation, and structured report generation via EchoPrime
GradCAM interpretability	Per-frame saliency overlays showing which image regions drove model predictions
DICOM viewer	Cornerstone.js-based playback with speed control, frame inversion, and overlay rendering
Draft report editing	Clinician reviews and edits AI-generated text before finalisation
Report finalisation	Irreversible locking enforced at both application layer (DbC) and database layer (trigger)
Word export	Finalised reports exported as .docx files for clinical documentation
Ethics consent gate	Mandatory consent form before any study functionality is accessible

5.5 Technology Stack

Table 3. Technology stack.

nublu		
Frontend	React, TypeScript, Vite	Browser-based clinical interface and SPA routing
DICOM viewer	Cornerstone.js	Frame playback, inversion, and canvas-based overlay rendering
Main backend	Node.js + Express (TypeScript)	Auth, patient records, studies, report lifecycle
Database	PostgreSQL	Persistent storage with integrity constraints and triggers
ML backend	Python, FastAPI	EchoPrime inference and GradCAM serving
Video encoder	MViT v2	Per-clip 512-dimensional video embeddings
View classifier	ConvNeXt	Echocardiographic view labelling
Text encoder	BioMedBERT	Biomedical text modality for contrastive training
Hosting backend	— Render	Free-tier Node.js service and PostgreSQL instance
Hosting — ML	Hugging Face Spaces	GPU-accessible FastAPI inference service

6. Project Execution

6.1 Project Timeline Across Two Semesters

An iterative approach over two semesters:

Semester 1 (Fall 2025). Research phase: understanding the clinical problem, reviewing existing literature, and evaluating possible models. Multiple pipelines got explored, including EchoCV and PanEcho. Finally, the semester ended with a decision to adopt EchoPrime for the final system.

Semester 2 (Spring 2026). Implementation phase: model integration, web platform development, usability evaluation. Key deliverables included the AI-backend-to-platform connection, study and report workflow, GradCAM interpretability, expert usability testing. By the semester end, the complete platform (upload, analysis, review, report finalisation) was operational.

6.2 Major Design Decisions

Full clinical decision-support platform. The scope expanded to a complete web platform, requiring frontend, backend, database, reporting workflow in addition to the models. This decision prioritised clinical usability.

Three-service microservice architecture. Separating frontend, application backend and ML backend allowed medical image processing to be isolated from user management, enabling DICOM data to be minimised at the application tier.

GradCAM interpretability. Adding spatial attribution via GradCAM was seen as essential for clinician trust: model outputs alone are not enough if physicians cannot verify the reasoning behind them.

Immutable report finalisation. Editable drafts as well as irreversible finalisation reduces the risk of accidental modification after sign off, a critical property for medical records.

6.3 Changes in Project Scope or Direction

The most important change was the changeing from PanEcho to EchoPrime. PanEcho's view-agnostic design produced in-study measurement variance of 10–30%, which is clinically unacceptable; EchoPrime's view-informed attention mechanism fixed this.

The project scope also expanded to a full web platform with user workflows, database-backed study management, draft editing and report export.

6.4 Challenges Encountered

Model selection and validation. Echocardiography analysis needs video data, multiple cardiac views and clinical interpretation, making model validation very demanding. Quite some models were checked before EchoPrime was chosen.

Deployment and hosting. Free-tier hosting on Render does not provide GPU access, creating a need for a separate Hugging Face Spaces deployment for ML inference.

Interpretability. The EchoPrime did not allow to show spatial reasoning to clinicians. A custom GradCAM module was added to close this gap.

Data and domain constraints. No local Kazakhstani echocardiographic dataset was available, restricting fine-tuning and imposing strict protocols on how studies could be uploaded and handled.

6.5 Solutions and Problem Handling

- **Modelling uncertainty:** Iterative exploration with structured comparison. PanE-cho rejected after validation showed higher than needed variance. EchoPrime adopted.
- **Architectural complexity:** Three-service decomposition isolated failures and bottlenecks.
- **Explainability:** Custom GradCAM pipeline implemented on top of EchoPrime's MViT v2 encoder.
- **Usability:** Expert sonographer demonstration collected structured feedback, three features added in response (GradCAM toggle, editable report fields, Word export).

6.6 Teamwork and Collaboration

Both semesters the project was divided into two sub-teams working on complementary components.

Semester 1. Team A: computer vision pipeline (view classification, measurement extraction).

Team B: NLP pipeline (tabular-to-text generation, ICD-10 classification).

Semester 2. One group implemented GradCAM on EchoPrime, the other built the frontend, backend, and hosting infrastructure.

Two weekly meetings were held throughout: the first to set weekly goals, the second to review progress, discuss blockers, brainstorm solutions. GitHub was used for task tracking and version control.

7. Evaluation

7.1 Evaluation Objectives

Three questions had to be answered for the evaluation of the project and its objectives:

1. Does platform streamline the clinicians reviewing process?
2. Are the predictions made by AI precise enough to reduce misinterpretation of cardiac measurements?
3. Does the system produce sufficiently transparent and practical outputs for clinicians work?

7.2 Evaluation Methodology

Two main methods were:

Weekly CTO meetings. The team held weekly meetings with the CTO of CheckApp, a digital-health startup. During each session CTO was updated on the web-application progress and gave feedback on product development, model improvements, and UI/UX for ease of use.

Expert product review. Sonographer with specialisation in echocardiography was invited for the review of the product. During the session, she evaluated the state of the product and fit for the established clinical workflow. She highlighted positive sides, and identified gaps that were further addressed by the team.

7.3 Validation of the Proposed Solution

Objective 1: Reducing specialist dependency. The platform shows that structured echo reports can be generated automatically from DICOM input without requiring an expert in the initial analysis step, making it usable in specialist settings that lack specialists.

Objective 2: Measurement accuracy. EchoPrime reached a mean AUC of 0.85–0.92 across five hospitals in three countries, outperformed prior foundation models, matched task-specific models and in some comparisons exceeded human cardiologist consistency (reported accuracy 0.89 vs 0.82).

Objective 3: Transparency and usability. Expert demonstration showed that GradCAM overlays, editable report fields and Word export together make the system practical for clinical workflow integration.

7.4 Discussion of Findings

Overall, evaluation results show that Zhurek AI is a credible, functional prototype for AI-assisted echocardiography interpretation. Strong model metrics and usability feedback support the claim that the system addresses the three objectives shown earlier.

However, appropriate caution is needed: strong metric performance does not guarantee readiness for real clinical deployment. Identified limitations include hardware demands, funding constraints, lack of local Kazakhstani fine-tuning data, and incomplete ICD-10 integration. The current system should be viewed as a promising prototype rather than a finished medical product.

The most important insight from evaluation is that the platform must be not only reliable but also interpretable and editable: GradCAM, editable reports and export capability were among the most valued features identified by the expert user, showing that clinician trust and workflow compatibility are the early requirements for adoption.

8. Conclusion and Future Work

8.1 Conclusion

Zhurek AI shows that an AI-based echocardiography decision-support system can be built and deployed in the scope of a student project using state-of-the-art vision-language models and domain-driven software architecture.

The system meets the main goals:

- Automates cardiac view classification and report generation relatively reliably.
- Provides explainable AI output using per-frame GradCAM attribution.
- Makes sure all clinical decisions remain under physician control.

Design decisions were graded by a practicing professional, and three features were added in response to the feedback.

Key changes include:

1. A three-service cloud architecture that decouples patient metadata from imaging data (privacy by design).
2. A domain-driven aggregate-root pattern enforcing clinical state machines at both application and database levels.
3. A custom GradCAM implementation for MViT v2 that allows spatial interpretability to embedding-based video models.

8.2 Future Work

Local fine-tuning. The next step is fine-tuning EchoPrime on a labelled Kazakhstani echocardiographic dataset. This requires an institutional data-sharing agreement and GPU budget.

Broader clinical validation. A multi-site prospective study involving cardiologists, echocardiographers, and general physicians across urban and rural settings would determine whether the platform reduces time-to-diagnosis and improves clinical outcomes in the population.

ICD-10 integration. Automated mapping to [ICD-10](#) diagnostic codes and integration with Kazakhstan's national health information system would allow effective epidemiological monitoring.

Federated learning. Applying [federated learning](#) would allow the model to improve across hospitals without centralising raw patient data, balancing model improvement with data privacy.

Mobile and low-bandwidth deployment. Optimising the inference pipeline for lower-spec hardware would extend the platform to rural and district hospitals, directly solving the health-equity gap identified in the constraints section.

9. Ethical and Legal Considerations

9.1 Identification of Ethical Issues

- **Patient privacy and confidentiality.** All data used in development is fully de-identified. No attempt is made to re-identify patients.
- **Responsible use of clinical data.** Datasets are used exclusively for research, in accordance with applicable data-use agreements.
- **Risk of misinterpretation.** Generated notes are clearly labelled as AI drafts. Phrasing is reviewed to avoid implying medical certainty.
- **Overreliance.** Each generated report is accompanied by a confidence indicator and requires physician sign-off before clinical use.

9.2 Legal Compliance

All team members completed CITI Data or Specimens Only Research training prior to accessing MIMIC datasets on PhysioNet. External libraries were used under compliant open-source licences. Patient data samples obtained from the university medical centre were handled strictly in accordance with CITI regulations.

9.3 Sustainability and Social Responsibility

The system is designed as a *decision-support tool*, not a replacement for clinical expertise. GradCAM heatmaps and editable report fields are intended to keep the physician in the loop at every step. Future deployment will prioritise equitable access, including optimisation for rural infrastructure.

References

- [1] Ministry of Health of the Republic of Kazakhstan. *Order approving the standard of cardiac and cardiac surgical care* (2022). <https://adilet.zan.kz/rus/docs/V2100026401>
- [2] M. Vukadinovic et al., “EchoPrime: A multi-video view-informed vision-language model for comprehensive echocardiography interpretation,” *Nature*, vol. 650, pp. 970–979, 2026. [doi:10.1038/s41586-025-09850-x](https://doi.org/10.1038/s41586-025-09850-x)
- [3] A. Markabayeva, A. S. Kerimkulova, R. Nurpeissova, and A. Ospanova, “Prevalence of cardiovascular diseases in Kazakhstan for 2013–2020,” *Journal of Hypertension*, vol. 42, suppl. 1, p. e162, 2024. [doi:10.1097/01.hjh.0001021004.41104.0b](https://doi.org/10.1097/01.hjh.0001021004.41104.0b)
- [4] A. Issanov et al., “Heart failure mortality trends in Kazakhstan 2014–2019: analysis of the Unified National Electronic Healthcare System.” [Internal citation based on project materials.]
- [5] A. Akanov et al., “Mortality trends in Kazakhstan: insights from a million deaths from 2014 to 2022,” *BMC Public Health*, 2025. [doi:10.1186/s12889-025-22225](https://doi.org/10.1186/s12889-025-22225)
- [6] C. R. Hoffmann et al., “Time requirements of the standard echocardiogram,” *Journal of the American Society of Echocardiography*, 2003. [doi:10.1016/S0894-7317\(03\)00590-3](https://doi.org/10.1016/S0894-7317(03)00590-3)
- [7] British Society of Echocardiography, “A practical guideline for performing a comprehensive transthoracic echocardiogram in adults,” *Echo Research and Practice*, 2021. [doi:10.1530/ERP-20-0026](https://doi.org/10.1530/ERP-20-0026)
- [8] B. Chong et al., “Global burden of cardiovascular diseases: projections from 2025 to 2050,” *European Journal of Preventive Cardiology*, vol. 32, no. 11, pp. 1001–1015, 2025.
- [9] T. H. Marwick, “The role of echocardiography in heart failure,” *Journal of Nuclear Medicine*, vol. 56, suppl. 4, pp. 31S–38S, 2015.
- [10] J. Marangou et al., “Echocardiography in indigenous populations and resource-poor settings,” *Heart, Lung and Circulation*, vol. 28, no. 9, pp. 1427–1435, 2019.
- [11] A. Radford et al., “Learning transferable visual models from natural language supervision,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 8748–8763.
- [12] J. Zhang et al., “Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy,” *Circulation*, vol. 138, no. 16, pp. 1623–1635, 2018. [doi:10.1161/circulationaha.118.034338](https://doi.org/10.1161/circulationaha.118.034338)

-
- [13] M. Christensen, M. Vukadinovic, N. Yuan, and D. Ouyang, “Vision–language foundation model for echocardiogram interpretation,” *Nature Medicine*, vol. 30, pp. 1481–1488, 2024.
 - [14] G. Holste, E. K. Oikonomou, M. Tokodi, A. Kovács, Z. Wang, and R. Khera, “Complete AI-enabled echocardiography interpretation with multitask deep learning,” *JAMA*, vol. 334, no. 4, pp. 306–318, 2025.
 - [15] Us2.ai, “Us2.ai receives FDA clearance for Us2.v2, transforming cardiovascular ultrasound analysis,” Press Release, Apr. 2024.
 - [16] Ultramics, “EchoGo Core 2.0: Next generation AI platform offers faster, more robust, holistic echo diagnostics,” Press Release, 2023.
 - [17] Ultramics, “10 Years in the Making: The clinical application of AI-based echocardiography diagnostics,” Blog, Feb. 2024.

A. Glossary

EchoPrime	A large pre-trained model linking echocardiographic video to clinical language, trained on millions of paired echo clips and cardiologist reports.
CLIP	A training strategy that teaches the model to match a video with its correct report by pulling matched pairs together and pushing mismatched pairs apart in a shared numerical space.
GradCAM	An explainability technique tracing model predictions back through network layers to highlight which pixels in which frames most influenced the output.
MViT v2	A video transformer that processes all frames of an echo clip simultaneously and compresses them into a single compact numerical representation.
BioMedBERT	A language model pre-trained on medical text that converts written cardiologist reports into numerical vectors comparable to video representations.
ConvNeXt	An image classifier identifying which standard cardiac view each echo frame depicts before the model interprets what it sees.
MIL	<i>Multiple Instance Learning</i> — a framework weighting each echocardiographic view by its relevance to the cardiac structure being assessed.
RAG	<i>Retrieval-Augmented Interpretation</i> — an inference strategy where the model retrieves the most similar historical reports from a database to guide new report generation.
Federated Learning	A privacy-preserving training approach where hospitals improve a shared model by contributing only learned parameter updates, never raw patient data.
DICOM	<i>Digital Imaging and Communications in Medicine</i> — the international standard file format for medical imaging data.
LVEF	<i>Left Ventricular Ejection Fraction</i> — the percentage of blood pumped out of the left ventricle with each heartbeat; a key indicator of cardiac function.
HFpEF	<i>Heart Failure with Preserved Ejection Fraction</i> — heart failure in which the ejection fraction appears normal but the heart fails to fill properly between beats.

ICD-10 *International Classification of Diseases, 10th Revision* — the WHO standard for encoding diagnoses as alphanumeric codes.