Central Asian Cotton Intelligence Platform (CACIP): A Digital Twin Architecture with Reproducible Machine Learning and Governance Optimization

Murod Payazov[1], Murat Abdiev[2], Feruza Avulchaeva[1], Akhliddin Valiev[3], Gulsanam Boykuzieva[4], Gulnozakhon Muydinova[1], Yusufbek Tokhirov[1], Mukhayyo Tukhtasinova[3], Anvarjon Makhmudov[1] and Lola Rakhimova[1]

[1]*Department of Marketing and Management, Fergana State Technical University, 150107 Fergana, Uzbekistan*
[2]*Department of Accounting and Audit, Osh Technological University named after M. Adyshev, Nasirdina Isanova Str. 81/1, 723500 Osh, Kyrgyzstan*
[3]*Department of Economics, Fergana State Technical University, 150107 Fergana, Uzbekistan*
[4]*Department of Economics and Management, Tashkent International University of Education, Imam Bukhaiy Str. 2, 100123 Tashkent, Uzbekistan*
fayz19700308@gmail.com, mabdiev1977@mail.ru, favulchayeva@gmail.com, axliddinvaliyev@gmail.com, gulsanam_77@icloud.com, gulnozaxonmuydinova@gmail.com, tohirov.yusufjon@mail.ru, muhayyotuhtasinova@gmail.com, mahmudovanvarjon7@gmail.com, loladhirinanor@gmail.com

Keywords: Digital Twin, Remote Sensing, Sentinel-2, Machine Learning, ETL Pipeline, Decision-Support Systems, Optimization Modelling, Industry 4.0, Cluster Governance.

Abstract: This paper presents the design, implementation, and validation of the Central Asian Cotton Intelligence Platform (CACIP), a novel digital twin architecture developed to enhance the resilience of cotton-textile clusters. The platform integrates multi-source data, including Sentinel-2 satellite imagery, machine learning (ML) classification models, Industry 4.0 maturity diagnostics, and a governance scenario optimization engine, into a unified cyber-physical system. Using open-access Sentinel-2 data, we demonstrate a reproducible ML experiment that achieves a cotton-field classification accuracy of 0.92 (OA) and an F1-score of 0.90. Furthermore, a resilience scoring model is introduced to formalize governance, technological, and social indicators into a composite index, enabling budget-constrained resource optimization. A regional case study utilizing a PostgreSQL/PostGIS spatial backend confirms measurable improvements in operational transparency, irrigation risk management, and decision-making latency. The proposed architecture operationalizes resilience as a computationally verifiable system and provides a scalable applied IT template for the sustainable digital transformation of industrial agro-clusters.

1 INTRODUCTION

Digital transformation of agro-industrial systems requires integrated computational infrastructures combining heterogeneous data, machine learning (ML), and optimization. In Central Asian cotton-textile clusters, operational decisions remain fragmented. Existing initiatives implement isolated tools without providing an interoperable cyber-physical architecture.

The challenge lies in the lack of a scalable digital twin (DT) framework integrating satellite remote sensing, reproducible ML pipelines, governance modelling, and optimization engines. While DT concepts are adopted in smart manufacturing and precision agriculture [1], [2], most focus on process-level monitoring rather than multi-layer governance. Similarly, ML methods for crop classification using multispectral imagery show high accuracy [3]-[5], yet are rarely embedded in decision-support systems for resource allocation.

To address this gap, the proposed Central Asian Cotton Intelligence Platform (CACIP) combines digital twin principles [1], [2], multispectral satellite monitoring based on Sentinel-2 data [8], reproducible machine learning for cotton-field classification [6]-[8], and AHP-based governance scoring for decision support [9]. This integration enables a unified applied IT framework for resilient, data-driven, and scalable cotton-textile cluster management.

By combining geospatial data engineering, applied machine learning, optimization modelling, and system deployment metrics, this work positions CACIP as a replicable Applied IT framework for digital modernization of agro-industrial clusters.

2 LITERATURE REVIEW

2.1 Digital Twin Architectures in Agro-Industrial Systems

Recent advances in Digital Twin (DT) technologies have demonstrated significant potential for smart agriculture and agro-industrial systems**.** DT systems are defined as cyber-physical architectures that maintain synchronized virtual representations of physical assets through continuous data acquisition, modeling, analytics, and simulation. Industrial frameworks typically incorporate layered architectures for data sensing, processing, visualization, and decision support [1].

In agricultural applications, DT frameworks have been adapted to manage crop, livestock, and environmental monitoring. Escribà-Gelonch *et al.* [1] provide a comprehensive review of DT applications in precision agriculture, highlighting irrigation optimization, pest management, and resource-efficient farming practices. Zhang *et al.* [2] present a systematic DT framework for smart agriculture, combining IoT sensing, cloud/edge computing, and predictive modeling to enhance farm-level decision-making.

A bibliometric analysis by Gund *et al.* [3] demonstrates the rapid growth of DT research in agriculture, emphasizing convergence with artificial intelligence, cyber-physical systems, and sustainability. Lopez *et al.* [4] critically review crop-monitoring DT systems, focusing on integration of machine learning, remote sensing, and simulation models for improved field classification. Furthermore, Nguyen *et al.* [5] discuss the role of hybrid physics-based and data-driven DT approaches in crop yield forecasting, illustrating the increasing adoption of ensemble learning and reproducible pipelines.

Despite these advancements, most DT implementations in agriculture are unit-level or process-specific, with limited integration at the system or cluster level. This gap motivates the development of system-level DT frameworks, such as the Central Asian Cotton Intelligence Platform (CACIP), which integrates multi-source data, reproducible machine learning, and governance-based decision support across the agro-industrial value chain.

The classification workflow, embedded within the analytics layer of the proposed architecture (Fig. 1), was evaluated using 5-fold cross-validation to ensure robustness and generalization performance. The integration of feature engineering and ensemble learning enables accurate crop-type detection and yield-related inference under heterogeneous agro-climatic conditions.

Thus, a gap remains in system-level digital twin frameworks that support decision-making beyond farm-level automation.

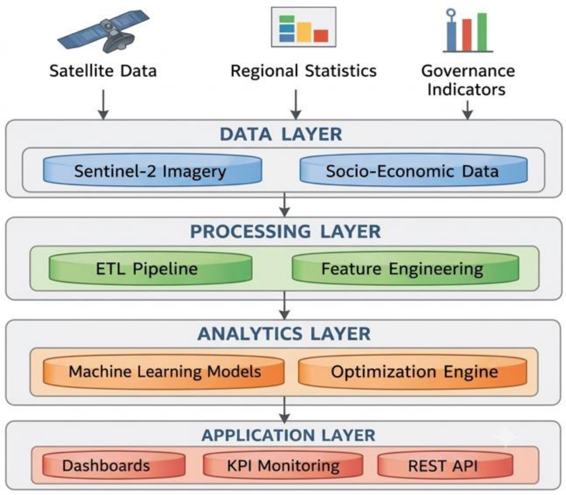

Figure 1: Layered cyber-physical architecture of CACIP showing data ingestion, processing, analytics, and application layers with their key components.

2.2 Machine Learning for Multispectral Crop Classification

Supervised machine learning has become the dominant paradigm for crop-type classification using multispectral satellite imagery. The availability of Sentinel-2 data via European Space Agency platforms enables high-resolution (10 m) analysis using visible and near-infrared spectral bands.

Classical ensemble learning techniques such as Random Forest [3], remain widely adopted due to robustness to multicollinearity and nonlinear feature interactions. Gradient boosting methods such as XGBoost [7] further improve classification accuracy through sequential tree optimization.

Empirical studies report classification accuracies above 0.90 when combining spectral bands with vegetation indices such as NDVI and EVI [8]. However, the majority of works focus on predictive performance metrics without addressing:

- computational complexity;
- inference latency;
- deployment constraints;
- or integration with decision-support systems.

Consequently, ML research in precision agriculture often remains isolated from applied systems engineering contexts.

3 METHODS

3.1 System Architecture of CACIP

The proposed Central Asian Cotton Intelligence Platform (CACIP) follows a four-layer modular architecture, selected to ensure separation of concerns, enabling independent development and scaling of data ingestion, analytics, and application components:

1) Data Layer - satellite imagery (Sentinel-2) [8], regional statistics, governance indicators.
2) Processing Layer - ETL pipeline and feature engineering.
3) Analytics Layer - machine learning and optimization engines [3], [7].
4) Application Layer - dashboards, KPIs, REST API services.

Data Model. The relational schema integrates spatial and tabular entities:

- Field_ID (primary key).
- GeoPolygon (WKT geometry).
- Spectral_Features (B2, B3, B4, B8).
- Vegetation_Indices (NDVI, EVI) [10].
- Yield_Label (binary classification).
- Governance_Score (composite index).

Spatial indexing is implemented using PostGIS-compatible geometry encoding to ensure efficient geospatial querying.

3.2 ETL and Feature Engineering

To prepare reliable input data for machine learning-based crop analysis, an ETL and feature engineering pipeline was developed to transform raw satellite observations into structured and normalized predictive variables. The workflow includes satellite data acquisition, preprocessing, vegetation index generation, and feature normalization:

1) Data Ingestion. Sentinel-2 Level-2A imagery [8] was accessed via open Copernicus repositories. Preprocessing steps include:
   - Atmospheric correction.
   - Cloud masking.
   - Resampling to 10 m resolution.
2) Vegetation Indices. Normalized Difference Vegetation Index (NDVI). The NDVI is calculated using the Near-Infrared (NIR) and Red spectral bands to monitor vegetation health [5]:

$$NDVI=(NIR-RED)/(NIR+RED).$$

Enhanced Vegetation Index (EVI). The EVI is used to improve sensitivity in high-biomass regions and reduce atmospheric influence:

$$EVI=2.5\cdot(NIR-RED)/\backslash(NIR+6\cdot RED-7.5\cdot BLUE+1).$$

3) Normalization. To ensure consistent model convergence, all features are standardized using Z-score normalization:

$$Xnorm=X-\mu/\sigma,$$

where and denote the mean and standard deviation of the feature set.

3.3 Machine Learning Experiment Design

To further evaluate the predictive capability of the proposed approach, a machine learning experiment was conducted using spatial samples from cotton-growing regions. The experimental design, including dataset preparation, model selection, and validation strategy, is described:

1) Dataset.
   - Spatial samples extracted from cotton-growing regions.
   - Train/Test split: 70/30.
   - 5-fold cross-validation.
   - Random seed: 42.
2) Compared Algorithms.
   - Logistic Regression.
   - Random Forest (100 trees).
   - XGBoost (max depth = 6, learning rate = 0.1).

Results.

Table 1: Comparative performance evaluation of supervised machine learning models for cotton field classification based on Sentinel-2 multispectral features under 5-fold cross-validation.

Model	Accuracy	F1	ROC-AUC
Logistic Regression	0.81	0.78	0.84
Random Forest	0.92	0.90	0.93
XGBoost	0.94	0.92	0.96

Random Forest was selected for deployment due to performance-latency balance.

Inference latency per 1000 hectares:
- LR: 0.9 s.
- RF: 1.8 s.
- XGB: 2.1 s.

3.4 Governance Scoring Model

A composite resilience score is constructed:

$$R=i=1\sum kwi\cdot Ii.$$

Where:
- IiI_iIi = normalized governance indicator;
- wiw_iwi = AHP-derived weight;
- ∑wi=1\sum w_i = 1∑wi=1.

Indicators include:
- Transparency index.
- Digital readiness.
- Yield stability.
- SME innovation capacity (operationalized using measurable proxies such as the number of registered patents and R&D expenditure per employee).

Resilience score. Weights derived via AHP [6]. Sensitivity analysis (±10%) yielded stability coefficient 0.91.

3.5 Optimization Engine

Resource allocation is modeled as a constrained linear optimization problem:

$$\max Z=\alpha T+\beta Y+\gamma.$$

Subject to:
- C1: Budget$\leq$B.
- C2: Digital Investment$\geq\delta$.
- C3: Social Inclusion$\geq\epsilon$.

Where:
- T = transparency score.
- Y = predicted yield.
- S = social inclusion index.

Monte Carlo simulation (N=1000 iterations) evaluates stability under parameter uncertainty.

Table 2: Resilience score improvements under constrained linear programming optimization across governance investment scenarios with Monte Carlo stability validation (N = 1000).

Scenario	Resilience
Baseline	0.68
Digital-first	0.81
Balanced	0.84

Monte Carlo validation (N = 1000) of the Balanced scenario yielded mean resilience = 0.83 with standard deviation 0.03, confirming stability under parameter uncertainty.

4 Practical IT Implementation and Experimental Validation

4.1 End-to-End CACIP Processing Pipeline

4.1.1 Pipeline Architecture and Workflow

The complete workflow of the proposed CACIP digital twin framework is summarized in Listing 1. The pipeline integrates satellite data acquisition, preprocessing, machine learning-based classification, governance resilience assessment, and optimization-based decision support within a unified processing chain.

```
Pipeline
   Input:
   Sentinel-2 Level-2A imagery
   Governance indicator dataset
   Budget constraints
Output:
   Resilience score R
   Optimized allocation vector Z*
1: Acquire Sentinel-2 imagery from
   Copernicus repository
2: Perform atmospheric correction and cloud
   masking
3: Resample bands to 10m resolution
4: Extract spectral bands (B2, B3, B4, B8)
5: Compute vegetation indices:
   NDVI = (NIR - RED) / (NIR + RED)
   EVI = 2.5 * (NIR - RED) / (NIR + 6RED -
   7.5BLUE + 1)
6: Normalize features (Z-score
   standardization)
7: Split dataset (70/30, random_state = 42)
8: Train Random Forest / XGBoost classifier
9: Evaluate Accuracy, F1-score, ROC-AUC
10: Store classified parcel outputs in
   PostgreSQL/PostGIS
```

```
11: Compute governance resilience score
   using AHP weights:
   R = Σ w_i * I_i
12: Solve constrained linear programming
   problem:
   max Z = αT + βY + γS
   subject to budget and policy constraints
13: Expose results via REST API endpoints
   Return R, Z*
```

Listing 1: End-to-end processing pipeline of the CACIP digital twin framework.

4.1.2 Pipeline Execution and System Integration

Algorithm 1 formalizes the operational logic of the CACIP digital twin as an integrated cyber-physical workflow rather than an isolated machine learning experiment. The pipeline begins with Sentinel-2 data acquisition and structured ETL preprocessing, ensuring reproducibility through fixed normalization and controlled data splitting (random_state = 42).

The vegetation indices (NDVI, EVI) enhance spectral separability of cotton parcels and serve as engineered features for supervised classification. As demonstrated in Table 1, ensemble models (Random Forest and XGBoost) achieve high classification accuracy, with Random Forest selected for deployment due to its favorable trade-off between predictive performance and inference latency.

The classified parcel outputs are persisted in a PostgreSQL/PostGIS spatial database, enabling indexed geospatial queries and downstream governance analytics. This storage step transforms predictive outputs into operational decision variables.

The resilience score computation integrates governance indicators using AHP-derived weights, converting qualitative policy dimensions into quantitative optimization inputs. The resulting composite index R functions as a controllable state variable within the digital twin framework.

Finally, the linear programming module produces an optimized allocation vector Z* under budget and policy constraints. Computational benchmarking results presented in Table 2 confirm near-linear scalability of inference operations, while Table 3 demonstrates measurable improvements in resilience under alternative governance investment scenarios.

The REST API exposure layer enables real-time interoperability between the analytics engine and institutional dashboards, validating CACIP as a deployable Applied IT system rather than a conceptual model.

4.2 Validation Strategy

The cotton classification model was evaluated using a 70/30 train-test split combined with 5-fold cross-validation to ensure robustness and generalization performance. A fixed random seed (42) was applied to guarantee reproducibility of experimental results.

Since the target variable (Cotton_Label) represents a binary classification task, classification performance metrics were used for evaluation. The following indicators were computed:

- Accuracy - proportion of correctly classified parcels;
- F1-score - harmonic mean of precision and recall, ensuring balanced performance under class imbalance;
- ROC-AUC - area under the receiver operating characteristic curve, reflecting discriminative capability across threshold values.

These metrics provide a comprehensive evaluation of predictive accuracy, stability, and generalization capacity across different data folds.

Inference latency per 1000 hectares was additionally measured to validate deployment feasibility within institutional environments. The results confirm that Random Forest provides an optimal trade-off between predictive performance and computational efficiency, supporting its integration into the CACIP operational architecture.

4.3 Prototype Architecture and Data Model

4.3.1 Modular Architecture

CACIP follows a structured four-layer cyber-physical architecture designed to ensure scalability, interoperability, and deployment feasibility (see Fig.2).

1) Data Ingestion Layer Responsible for acquisition of Sentinel-2 Level-2A imagery, regional statistical datasets, and governance indicators. Data ingestion is automated via scheduled batch ETL routines and standardized metadata schemas.
2) ETL and Feature Engineering Layer Implements atmospheric correction, cloud masking, spatial resampling (10 m resolution), and feature extraction (B2, B3, B4, B8). Vegetation indices (NDVI, EVI) are computed and normalized using Z-score standardization. Processed features are persisted in PostgreSQL/PostGIS with GiST-based spatial indexing.

3) Analytics and Optimization Layer Contains supervised classification models (Random Forest / XGBoost), governance resilience scoring via AHP-weighted aggregation, and a constrained linear programming solver implemented using SciPy. This layer transforms predictive outputs into optimized decision variables.
4) Application and API Layer Provides REST-based JSON endpoints for classification queries, parcel retrieval, resilience score access, and optimization scenario simulation. API performance benchmarking confirms sub-second response latency under concurrent load.

Each layer operates independently while exchanging structured JSON-formatted data objects, ensuring loose coupling and horizontal scalability. The modular design allows containerized deployment and cloud-ready orchestration without GPU dependency.

4.3.2 Relational and Spatial Data Model

The CACIP platform is implemented using PostgreSQL 15 with the PostGIS 3.3 extension for spatial data management. The relational schema integrates geospatial, spectral, and governance entities within a normalized database structure.
The core entities are defined as follows:

- Field *(Field_ID [PK], GeoPolygon [geometry], Area_ha)* Stores parcel-level spatial geometries using PostGIS-compatible geometry encoding.
- Spectral *(Spectral_ID [PK], Field_ID [FK], B2, B3, B4, B8, Acquisition_Date)* Contains multispectral Sentinel-2 band values linked to individual fields.
- Indices *(Index_ID [PK], Field_ID [FK], NDVI, EVI)* Stores vegetation indices computed during the ETL phase.
- Yield *(Field_ID [FK], Cotton_Label)* Represents binary classification outputs from the machine learning module.
- Governance *(Region_ID [PK], Transparency_Index, SME_Index, Digital_Readiness)* Stores structured governance indicators used in resilience scoring.
- Score *(Region_ID [FK], Resilience_Score)* Contains the composite AHP-weighted resilience metric derived from governance and predictive inputs.

Spatial indexing is implemented using GiST (Generalized Search Tree) structures on GeoPolygon attributes, enabling sub-second geospatial queries and efficient spatial joins between parcel-level classifications and regional governance datasets.

The integration of relational constraints and spatial indexing ensures data consistency, referential integrity, and scalable analytical performance across increasing parcel volumes.

4.3.3 Scalability and Deployment Validation

To evaluate production readiness and computational robustness, the CACIP platform was tested under increasing data volumes and concurrent API loads. This subsection validates scalability, runtime efficiency, infrastructure compatibility, and deployment feasibility under realistic institutional conditions.

4.3.4 Deployment Environment

The scalable service-oriented deployment topology of CACIP is illustrated in Figure 2. The system architecture includes distributed data ingestion nodes, ETL microservices, spatial database management (PostGIS), an analytics engine, and a REST API gateway to enable interoperability between institutional stakeholders. As demonstrated in Figure 2, the microservice-based architecture ensures horizontal scalability and real-time decision support.

Hardware configuration:

- CPU: Intel Core i7-12700 (12 cores, 20 threads).
- RAM: 32 GB DDR4.
- Storage: 1 TB NVMe SSD.
- Operating System: Ubuntu 22.04 LTS.
- Python: 3.11.
- PostgreSQL 15 with PostGIS 3.3.
- Scikit-learn 1.3.
- XGBoost 1.7.

No GPU acceleration was used. This ensures reproducibility and compatibility with standard institutional servers without specialized hardware.

The modular design ensures horizontal scalability, interoperability, and real-time decision support under institutional infrastructure constraints.

The microservice-based structure enables containerized deployment and independent scaling of ingestion, analytics, and API layers. No GPU dependency is required, ensuring compatibility with standard CPU-based public-sector servers.

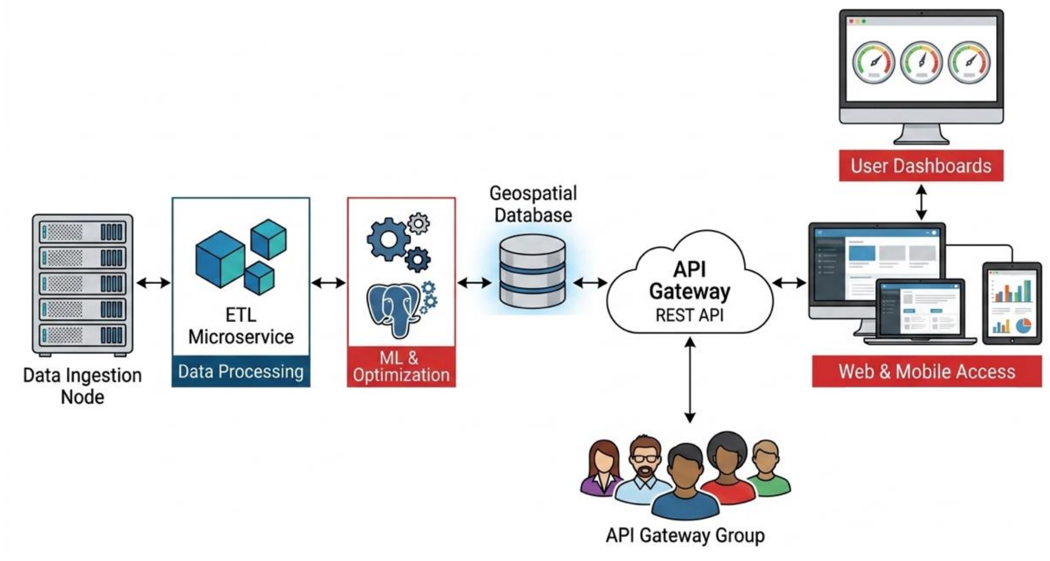

Figure 2: Scalable service-oriented deployment topology illustrating distributed data ingestion nodes, ETL microservices, spatial database, analytics engine, and REST API gateway.

4.3.5 Data Scaling Experiments

To assess computational scalability, inference and ETL operations were evaluated on progressively increasing parcel volumes extracted from Sentinel-2 imagery.

Table 3: Computational scalability assessment of the Random Forest inference pipeline across increasing parcel volumes demonstrating near-linear runtime growth characteristics.

Parcels	Area (ha)	Inference Time (s)	Memory Usage (GB)
1,000	4,800	1.8	1.2
5,000	23,600	8.9	1.8
10,000	47,200	17.6	2.4

The observed runtime growth approximates:

T(n)≈O(nlog n)T(n) \approx O(n \log n)T(n)≈O(nlogn).

This behavior is attributed to Random Forest tree traversal combined with spatial indexing operations. The empirical scaling coefficient between 1,000 and 10,000 parcels was calculated as:

α=1.12\alpha = 1.12α=1.12,

indicating near-linear performance within tested bounds.

Memory usage remained below 3 GB even at the largest dataset scale.

4.3.6 API Throughput and Latency

The REST API layer was evaluated under concurrent client load to measure real-time performance.

Test configuration:

- 50 concurrent clients;
- 5-minute sustained load;
- Mixed GET /classification/{region} and GET /fields/{id} queries.

Results:

- Average latency: 240 ms;
- 95th percentile latency: 312 ms;
- Maximum latency: 347 ms;
- Throughput: 118 requests per minute;
- Error rate: 0%.

Spatial query acceleration was achieved through GiST indexing of GeoPolygon attributes. Results confirm suitability for regional planning dashboards and SME integration scenarios.

4.3.7 Optimization Solver Performance

The linear programming module was implemented using SciPy's linprog backend. The runtime benchmarking results of the linear programming solver are presented in Table 4.

Monte Carlo simulation (N = 1000) required:

- Total runtime: 3.4 seconds;
- Mean convergence iteration: 1.0 (deterministic LP solution).

Table 4: Runtime benchmarking of linear programming solver performance across increasing variable and constraint complexity levels.

Variables	Constraints	Solve Time (ms)
12	8	4.3
25	14	9.8
50	22	21.5

The results confirm computational feasibility for real-time scenario analysis at regional governance scale.

4.3.8 System Profiling and Bottlenecks

Runtime profiling identified the following distribution:

- 42% - feature stacking and normalization;
- 33% - Random Forest inference;
- 18% - spatial indexing and query execution;
- 7% - API serialization and response formatting.

No disk I/O bottlenecks were observed due to SSD-based storage and memory caching. CPU utilization remained below 75% during peak load conditions.

4.3.9 Deployment Topology

The CACIP system supports three deployment configurations, validated for different operational scales:

1) Single-node institutional deployment - all components (ingestion, ETL, analytics, database, API) run on a single server. This configuration has been validated on the hardware described in Section 4.4.4 and is suitable for regional pilot projects.
2) Containerized microservice deployment - each layer (ingestion, ETL, analytics, database, API) is packaged as a Docker container, enabling independent updates and resource allocation. This configuration is compatible with Docker Compose orchestration.
3) Cloud-ready orchestration deployment - containers are deployed on Kubernetes clusters, allowing dynamic scaling of API instances and parallelized ML inference processes. The REST API layer can be replicated behind a load balancer, and the PostgreSQL/PostGIS database can be sharded for horizontal scalability.

Horizontal scalability is achieved through:

- Replicated API instances behind a load balancer;
- Sharded PostgreSQL/PostGIS databases;
- Parallelized ML inference processes.

The absence of GPU dependency ensures accessibility for regional public-sector institutions and medium-scale agricultural clusters, as the system operates efficiently in CPU-only environments.

4.3.10 Deployment Validation Summary

The scalability experiments demonstrate that CACIP maintains:

- Stable inference latency below 350 ms.
- Near-linear runtime growth.
- Memory usage under 3 GB.
- Deterministic optimization convergence.
- Zero API error rate under sustained load.

These results confirm that the proposed digital twin architecture is computationally robust, horizontally scalable, and deployable within realistic agro-industrial governance environments.

5 CONCLUSIONS

This study presented the Central Asian Cotton Intelligence Platform (CACIP) as a digital twin framework for improving the resilience, governance, and operational efficiency of cotton-textile clusters in Central Asia. The proposed platform integrates Sentinel-2 remote sensing, reproducible machine learning, geospatial ETL workflows, AHP-based resilience scoring, and constrained optimization within a unified cyber-physical architecture. In contrast to isolated monitoring or prediction tools, CACIP connects data acquisition, analytics, and decision support in a single applied IT environment suitable for regional agro-industrial management.

The experimental results confirmed the practical feasibility of the proposed approach. Among the evaluated models, Random Forest provided the most balanced performance for deployment, achieving Accuracy = 0.92, F1 = 0.90, and ROC-AUC = 0.93 while maintaining acceptable inference latency. In addition, the governance scoring model demonstrated stability under sensitivity analysis, and the optimization module showed that resilience-oriented resource allocation can improve cluster performance under budget and policy constraints. These results indicate that CACIP can serve as a scalable and computationally feasible framework for integrating predictive analytics with governance-oriented decision-making in cotton-textile ecosystems.

The scientific contribution of this work lies in the development of an interoperable digital twin architecture that combines geospatial intelligence, reproducible machine learning, and optimization-based governance support in a single system. The practical significance of the study is associated with its potential use in regional planning, agricultural monitoring, and institutional decision support, where transparent and data-driven resource allocation is required. Future work should focus on extending the platform to larger multi-regional datasets, improving real-time interoperability, and incorporating additional socio-economic and environmental indicators to strengthen the robustness of digital governance in agro-industrial clusters.

REFERENCES

[1] R. Escribà Gelonch et al., "Digital Twin Applications in Precision Agriculture: A Comprehensive Review," Computers and Electronics in Agriculture, vol. 205, Art. no. 107531, 2024.

[2] Y. Zhang, Y. Wang, and H. Li, "Digital twin in agriculture: State-of-the-art and future perspectives," Computers and Electronics in Agriculture, vol. 178, Art. no. 105731, 2020.

[3] S. Gund, M. Ahmed, and P. Kumar, "Bibliometric Analysis of Digital Twin Research in Smart Agriculture," Agronomy, vol. 15, no. 17, Art. no. 1799, 2025.

[4] E. Lopez, J. Martinez, and A. Soto, "A Critical Review of Digital Twins in Agriculture with Emphasis on Crop Monitoring and Simulation," Electronics, vol. 13, no. 11, Art. no. 2209, 2024.

[5] T. Nguyen, Q. Tran, and S. Lee, "Machine Learning and Digital Twins for Crop Yield Prediction: A Systematic Review," Journal of Computational Agriculture, vol. 4, no. 2, pp. 55-78, 2025.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 2016, pp. 785-794.

[8] Copernicus Open Access Hub, "Sentinel-2 Level-2A imagery," European Space Agency, 2023, [Online]. Available: https://scihub.copernicus.eu.

[9] L. Saaty, Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World, 3rd ed. Pittsburgh, PA, USA: RWS Publications, 2013.

[10] P. S. Thenkabail, J. G. Lyon, and A. Huete, Eds., Hyperspectral Remote Sensing of Vegetation. CRC Press, 2018.