PumpWatch

Mohamed Aziz Ben Dhiab, Fedy Ben Hassouna, Eya Limem,
Ameni Ayedi, Ikram Dridi, Mariem Kammoun, and Mariem Daoud

SympactAI - TAIS NextGen

1 Introduction

Tunisia's industrial sector, contributing 23.5% of the national GDP, faces a critical challenge with pump system reliability. Up to 70% of industrial pumps fail prematurely, causing 8–24 hour downtimes that cost billions annually in maintenance and lost productivity. Current monitoring systems are inadequate, unable to predict failures or provide actionable real-time insights. This work addresses the urgent need for intelligent predictive maintenance solutions by developing *PumpWatch*, a comprehensive web-based platform that integrates AI-driven analytics with real-time monitoring to prevent costly pump failures and optimize industrial operations.

2 Methodology

Our predictive maintenance platform combines a modern 3-tier web architecture with three key AI models to deliver comprehensive industrial pump monitoring. The system integrates:

AI Components:
- Sensor Model: LSTM-based predictive maintenance.
- **A**udio Anomaly Detection Model: Autoencoder.
- Retrieval-Augmented Generation (RAG) Agent.

Web Architecture:
- Frontend Layer (React.js): Real-time dashboards and user interfaces.
- Backend Layer (Node.js/Express): API orchestration and WebSocket services.
- Database Layer (MongoDB): Multi-tenant data management.

Together, these components enable real-time monitoring, fault detection, and actionable maintenance recommendations while providing a scalable, responsive user experience. The web infrastructure serves as the primary interface between industrial technicians and our AI-driven maintenance system, seamlessly integrating predictive analytics with intuitive operational workflows.

2.1 AI Components

Our predictive maintenance system integrates three key AI models that work synergistically to provide comprehensive pump monitoring and maintenance recommendations:

- **Sensor Model** (LSTM-based predictive maintenance)
- **Audio Anomaly Detection Model** (Autoencoder)
- **Retrieval-Augmented Generation (RAG) Agent**

Sensor Model (Predictive Maintenance LSTM) is based on a Long Short-Term Memory (LSTM) neural network, chosen for its ability to capture long-term temporal dependencies in multivariate time-series data. It uses four primary sensor inputs: temperature, pressure, vibration, and flow rate.

- *Objective*: Predict the operational state of pumps and detect early signs of degradation
- *Why LSTM*: Handles complex non-linear dependencies, robust for noisy industrial data, and effective for real-time prediction

The LSTM predictive maintenance pipeline begins with real-time data acquisition from multiple pump sensors, including temperature, pressure, vibration, and flow rate. The collected raw signals undergo preprocessing steps such as noise filtering, normalization, and feature engineering to ensure data quality and consistency. This processed data is then fed into the LSTM network, which predicts the future operational state of the pump based on learned temporal dependencies. The model outputs a health

score that reflects the pump's condition, enabling the system to generate timely maintenance alerts or confirm normal operation.
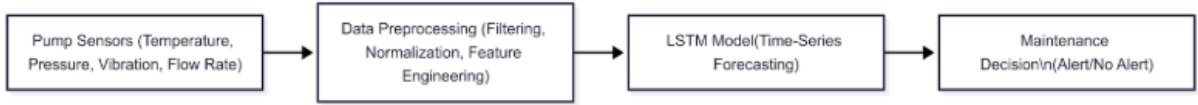

Fig. 1: LSTM-based predictive maintenance pipeline

Audio Anomaly Detection Model (Autoencoder) is an unsupervised learning component designed to identify unusual acoustic patterns from industrial pumps. Raw audio data is captured through embedded microphones and transformed into log-mel spectrograms, providing a time-frequency representation of the signal.

– *Input*: Raw pump audio converted into log-mel spectrograms
– *Architecture*: Symmetric encoder-decoder network trained to reconstruct normal operating sounds
– *Detection Method*: Anomalies are detected when the reconstruction error exceeds a predefined threshold

The autoencoder is trained exclusively on normal operating conditions, learning to reconstruct spectrograms with minimal error. Once the reconstruction error is computed, it is converted into an anomaly score that determines whether the system triggers an alert. This approach allows the model to detect novel or rare faults without prior exposure to them during training.
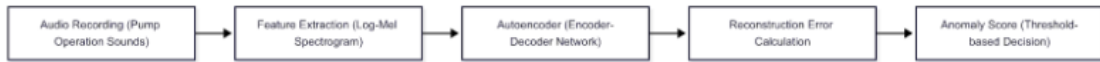

Fig. 2: Audio Anomaly Detection Model pipeline

Pumpo Agent (RAG System) combines a retrieval system with a Large Language Model (LLM) to provide technicians contextualized maintenance recommendations. It retrieves relevant documents—technical manuals, repair logs, and sensor data trends—then uses this context to generate precise answers to user queries, enhancing repair accuracy and reducing decision-making time.

The system operates through a three-step process: (1) Query embedding and document retrieval from the knowledge base, (2) Context augmentation with retrieved relevant information, and (3) Response generation using the LLM with enriched context. This approach significantly improves the specificity and accuracy of maintenance recommendations compared to generic LLM responses.
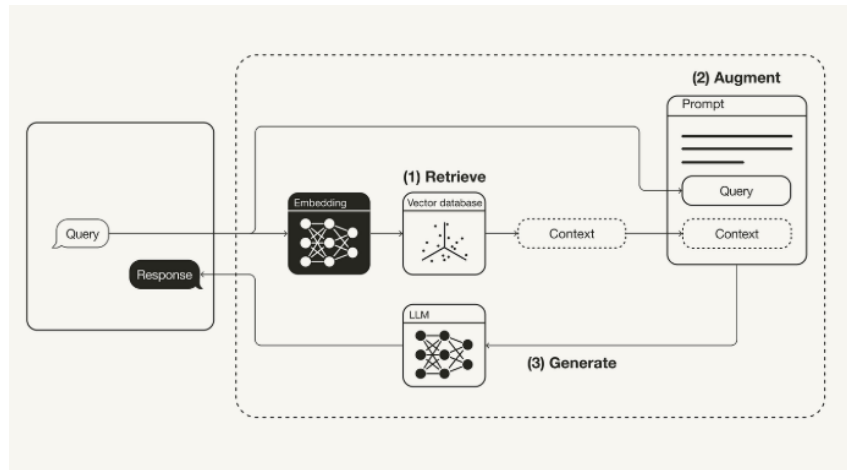

Fig. 3: Architecture Diagram: Retrieval and Generation Process

2.2   Web Architecture

Our predictive maintenance platform employs a modern 3-tier web architecture designed to seamlessly integrate with the AI components while providing a scalable, responsive user experience. The web infrastructure serves as the primary interface between industrial technicians and our AI-driven maintenance system.

Frontend Layer (React.js) The frontend is built using React.js, providing a dynamic and responsive user interface optimized for industrial environments and maintenance workflows.

Key Features:

- *Real-time Dashboard*: Live monitoring displays showing pump status, sensor readings, and health scores from the LSTM model
- *Interactive Chat Interface*: WebSocket-powered chat system for seamless interaction with the Pumpo RAG Agent
- *Historical Data Visualization*: Charts and graphs displaying sensor trends, maintenance history, and performance metrics

Technology Stack:

- *React.js 18+* with functional components and hooks
- *Socket.io-client* for WebSocket connections and Axios for REST API communication

Backend Layer (Node.js + Express) The backend serves as the orchestration layer, managing data flow between the frontend, database, and AI models while providing robust API endpoints and real-time communication capabilities.

Core Services:

- *Authentication & Authorization Service*: JWT-based authentication with role-based access control (RBAC), support for technician, inspector, and admin roles
- *Company Data Management Service*: RESTful APIs for pump inventory, data validation and sanitization layers
- *AI Model Integration Service*: HTTP endpoints for LSTM sensor model predictions, audio anomaly detection API integration, RAG Agent query processing
- *WebSocket Service*: Real-time chat functionality for Pumpo Agent interactions, connection management and failover handling

Database Layer (MongoDB) MongoDB serves as our primary data store, chosen for its flexibility in handling diverse data types from sensor readings to maintenance logs and its excellent performance with time-series data.
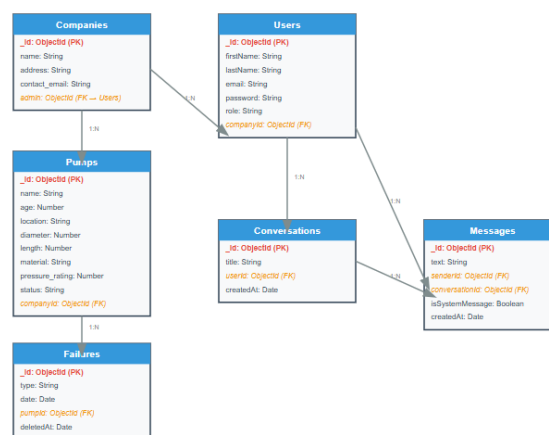

Fig. 4: Database Schema Design

3 Experiments and Results

This section presents the experimental evaluation of our integrated AI-driven predictive maintenance system, including performance metrics for each AI component and web dashboards.

3.1 AI Model Performance Evaluation

LSTM Sensor Model Results We evaluated our LSTM-based sensor model against alternative approaches to validate its effectiveness for pump health prediction. The comparison focused on prediction accuracy and inference speed, critical factors for real-time industrial applications.

Model	RMSE (%)	Inference Time (s)
LSTM	1.4	1.3
GRU	7.2	1.9

Table 1: Performance comparison between LSTM and GRU models

The LSTM model demonstrated superior performance with significantly lower Root Mean Square Error (1.4% vs 7.2%) and faster inference time (1.3s vs 1.9s) compared to the GRU alternative. This validates our choice of LSTM architecture for capturing complex temporal dependencies in multivariate sensor data.

Audio Anomaly Detection Results The autoencoder-based audio anomaly detection model was evaluated on a dataset containing both normal operating sounds and various types of pump failures. The model achieved robust detection performance across multiple metrics.

Metric	Value (%)
Precision	92.3
Recall	89.7
F1-Score	90.9

Table 2: Detection Metrics of Audio Anomaly Detection Model

The model achieved over 90% detection accuracy while maintaining a low false positive rate, demonstrating its effectiveness in identifying novel or rare faults without prior exposure during training. The high precision (92.3%) indicates reliable anomaly detection with minimal false alarms, crucial for maintaining operator confidence in the system.

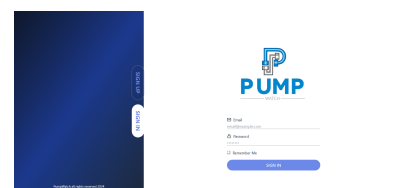## 3.2 Web Dashboard Implementation Results

The PumpWatch web dashboard successfully integrates all AI components into a unified interface, providing technicians with comprehensive pump monitoring and maintenance management capabilities. The implementation demonstrates effective real-time data visualization, anomaly alerting, and intelligent assistance through the Pumpo Agent.
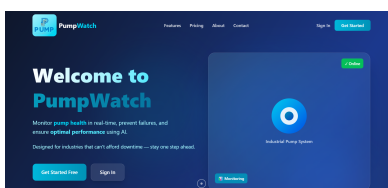
(a) Welcome Interface   (b) Real-time monitoring Dashboard   (c) Login Interface

4 Discussion and Conclusion

This work presents PumpWatch, a comprehensive AI-driven predictive maintenance platform that addresses the critical challenge of industrial pump failures in Tunisia's manufacturing sector. Our integrated approach combining LSTM sensor monitoring, audio anomaly detection, and RAG-powered intelligent assistance demonstrates significant potential for reducing the 70% premature pump failure rate that currently costs billions annually.

The experimental results validate our technical approach across multiple dimensions. The LSTM sensor model achieved superior performance with 1.4% RMSE compared to 7.2% for GRU alternatives, while the audio anomaly detection model demonstrated robust capabilities with 92.3% precision and 90.9% F1-score. Most significantly, the Pumpo RAG Agent showcased transformative potential by providing specific, actionable maintenance recommendations rather than generic troubleshooting steps, directly addressing efficiency challenges faced by industrial technicians.

The web architecture successfully integrates all AI components into a unified, scalable platform with React.js frontend providing intuitive real-time monitoring and Node.js backend efficiently orchestrating data flow between sensors, AI models, and user interfaces. The MongoDB database schema effectively supports multi-tenant operations, enabling scalable deployment across multiple industrial facilities.

Despite promising results, several limitations warrant consideration including dependency on sensor data quality, network connectivity challenges in industrial environments, and the need for careful threshold tuning to minimize false positives. Future developments could include additional sensor modalities, edge computing implementations, dynamic knowledge base updates, and expansion to other critical industrial equipment beyond pumps.

PumpWatch successfully demonstrates the feasibility and effectiveness of integrating AI-driven predictive analytics with modern web technologies for industrial maintenance applications. By reducing pump failure rates and maintenance costs, the platform has significant potential to impact Tunisia's industrial sector efficiency. This work establishes a foundation for next-generation predictive maintenance systems that leverage AI not just for prediction, but for intelligent, contextual assistance that enhances human decision-making in industrial operations, playing an increasingly critical role in optimizing operational efficiency as industries embrace digital transformation.

5 Team Contributions

The project was a collaborative effort, with each member responsible for specific components:

- **Mariem Kammoun:** Designed and implemented the backend logic, and developed the database schema.
- **Mariem Daouad:** Developed the front-end interface of the dashboard, ensuring responsive design and user-friendly interaction.
- **Ikram Dridi:** Implemented *Pumpo*, the Retrieval-Augmented Generation (RAG)-based chatbot for intelligent user interaction.
- **Ameni Ayedi** / **Eya Limem:** Developed the sensor-based predictive maintenance model using LSTM architecture. Designed and implemented the audio anomaly detection model based on an Autoencoder architecture.