

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Higher School of Computer Science
– 08 May 1945 – Sidi Bel Abbès



THESIS

To obtain the diploma of **State Engineer**

Field: **Computer Science**

Specialty: **Artificial Intelligence and Data Science (AIDS)**

Theme:

AI-Based Verification of Telecom Site Preventive Maintenance

From a Multi-Expert Vision Pipeline to a Unified Fine-Tuned Vision–Language Model

Presented by:

Saadi Ahmed

Carried out within **Ooredoo Algeria**

Submission Date: 29/06/2026

In front of the jury composed of:

Dr. KAZI TANI Mohammed Yassine

President

Dr. TAOULI Amina

Examiner

Dr. DIF Nassima

Supervisor

Mr. BERRADJA ihssene

Co-Supervisor

Academic Year: **2025 / 2026**

Abstract

Telecom operators rely on the continuous availability of thousands of geographically dispersed cell sites, whose reliability depends on periodic *preventive maintenance*. At Ooredoo Algeria, the maintenance carried out by subcontractors on roughly ten thousand sites is checked by a small verification team that manually inspects the photographic evidence of every checklist item, a slow, subjective and hardly scalable process.

This report presents the design, training and evaluation of an artificial-intelligence system that assists this visual verification. The work tackles a deliberately difficult real-world setting: the imagery had to be recovered indirectly, it is weakly labelled (about 86% of items carry no usable verdict, combining *submitted* and *initial* statuses; about 9% are *fail* verdicts; and about 4.8% are pass-type verdicts), and many decisive cues sit on blurry seven-segment instrument displays.

Two complementary approaches are developed. The first is a **multi-expert pipeline**, in which a specialised model is dedicated to each verification family: cleanliness, structural damage, connection and wiring, before/after comparison, leakage, component presence and identification, component counting and optical character recognition, combining convolutional classifiers, CLIP text-image embeddings, self-supervised DINOv2 features, multiple-instance learning and YOLO detection. The second is a **unified fine-tuned Vision–Language Model** (Qwen3.5-4B for fast iteration and Qwen2.5-VL-7B for the retained evaluated run, adapted with QLoRA on constrained GPU resources) that performs the checks through natural-language instructions and naturally extends the pipeline. The work is integrated into a full software platform: a Flutter mobile application for field agents, a React supervision dashboard, and a Node.js/Express backend connected to Ooredoo’s service-management system.

Keywords: Computer Vision, Vision–Language Models, Fine-Tuning (QLoRA), Optical Character Recognition, Weakly-Supervised Learning, Telecom Preventive Maintenance, Automated Visual Inspection, Ooredoo.

Résumé

Les opérateurs télécoms dépendent de la disponibilité permanente de milliers de sites radio dispersés géographiquement, dont la fiabilité repose sur une *maintenance préventive* périodique. Chez Ooredoo Algérie, la maintenance réalisée par des sous-traitants sur environ dix mille sites est contrôlée par une petite équipe de vérification qui inspecte manuellement les preuves photographiques de chaque point de contrôle, un processus lent, subjectif et difficilement extensible.

Ce mémoire présente la conception, l'entraînement et l'intégration d'un système d'intelligence artificielle assistant cette vérification visuelle. Le travail s'attaque à un cadre réel volontairement difficile : les images ont dû être récupérées indirectement, elles sont faiblement annotées (environ 86% des items ne portent aucun verdict exploitable, en combinant les statuts *soumis* et *initial*; environ 9% sont des verdicts *échoué*; et environ 4,8% sont des verdicts de type *succès*), et de nombreux indices décisifs se trouvent sur des afficheurs à sept segments peu lisibles.

Deux approches complémentaires sont développées. La première est une **chaîne multi-experts**, où un modèle spécialisé est dédié à chaque famille de vérification : propreté, dégradations structurelles, connectique et câblage, comparaison avant/après, fuites, présence et identification de composants, comptage de composants et reconnaissance optique de caractères, combinant classifieurs convolutionnels, embeddings texte-image CLIP, descripteurs auto-supervisés DINOv2, apprentissage à instances multiples et détection YOLO. La seconde est un **modèle vision-langage unifié et finétuné** (Qwen3.5-4B pour l'itération rapide et Qwen2.5-VL-7B pour le modèle retenu, adapté par QLoRA sur des ressources GPU limitées) qui réalise les contrôles via des instructions en langage naturel et prolonge naturellement la chaîne. Le travail est intégré dans une plateforme logicielle complète : une application mobile Flutter pour les agents de terrain, un tableau de bord de supervision React, et un backend Node.js/Express connecté au système de gestion de service d'Ooredoo.

Mots-clés : Vision par ordinateur, Modèles vision-langage, Fine-tuning (QLoRA), Reconnaissance optique de caractères, Apprentissage faiblement supervisé, Maintenance préventive télécom, Inspection visuelle automatisée, Ooredoo.

ملخص

يعتمد مشغلو الاتصالات على التوفر المستمر لآلاف المواقع الخلوية المنتشرة جغرافيًا، والتي تتوقف موثوقيتها على الصيانة الوقائية الدورية. في أوريدو الجزائر، تخضع أعمال الصيانة التي ينجزها متعاقدون من الباطن على نحو عشرة آلاف موقع للفحص من قبل فريق تحقق صغير يقوم يدويًا بفحص الأدلة الفوتوغرافية لكل عنصر في قائمة التحقق، وهي عملية بطيئة وذاتية وغير قابلة للتوسع بسهولة.

يقدم هذا التقرير تصميم وتدريب وتقييم نظام للذكاء الاصطناعي يساعد في هذا التحقق البصري. يتعامل هذا العمل مع سياق حقيقي صعب عمدًا: تعيين استعادة الصور بشكل غير مباشر، وهي ضعيفة التصنيف، حيث إن نحو 86% من العناصر لا تحمل حكمًا قابلاً للاستخدام، يجمع بين الحالات المرسله والابتدائية؛ ونحو 9% أحكام فشل؛ ونحو 4.8% أحكام من نوع النجاح. كما أن كثيرًا من المؤشرات الحاسمة توجد على شاشات عرض رقمية ذات سبع قطع (seven-segment) غير واضحة.

تم تطوير نهجين متكاملين. الأول هو خط أنابيب متعدد الخبراء (multi-expert)، حيث يخصص نموذج متخصص لكل فئة من فئات التحقق: النظافة، الأضرار البنيوية، التوصيلات والأسلاك، المقارنة قبل وبعد، التسرب، وجود المكونات وتحديداتها، عد المكونات، والتعرف الضوئي على الحروف (OCR)، بالجمع بين مصنفات الشبكات العصبية التلافيفية، وتمثيلات CLIP النصية البصرية، وخصائص DINOv2 ذات التعلم الذاتي الموجه، والتعلم متعدد النماذج (multiple-instance learning)، والكشف بواسطة YOLO.

والنهج الثاني هو نموذج لغوي بصري موحد (Vision-Language Model) مضبوط دقيقًا، وهو Qwen3.5-4B للترار السريع ونموذج Qwen2.5-VL-7B للتشغيل المعتمد والمقيم، ومكيف باستخدام QLoRA على موارد معالجة رسومية محدودة. ينفذ هذا النهج عمليات التحقق من خلال تعليمات باللغة الطبيعية ويوسع خط الأنابيب بشكل طبيعي. تم دمج هذا العمل في منصة برمجية كاملة: تطبيق هاتفي بتقنية Flutter للعملاء الميدانيين، ولوحة تحكم إشرافية بتقنية React، وخلفية (backend) بتقنية Node.js/Express متصلة بنظام إدارة الخدمات الخاص بأوريدو.

الكلمات المفتاحية: الرؤية الحاسوبية، النماذج اللغوية البصرية، الضبط الدقيق (QLoRA)، التعرف الضوئي على الحروف، التعلم ضعيف الإشراف، الصيانة الوقائية للاتصالات، الفحص البصري الآلي، أوريدو.