**Documentation**

**Contrôle qualité automatisé des vis**

**MADOUNGOU Colombe Alice**

**HACHICHA Soumaya**

**EBING'NE Cheryle**

**ELDRISSI Yasmine[[1]](#footnote-0)**

**1. Contexte Général**

L’entreprise, spécialisée dans la fabrication de vis, a décidé d’automatiser son contrôle qualité en utilisant un système de vision par ordinateur basé sur l’intelligence artificielle. L’objectif est d’identifier les défauts de fabrication afin d’améliorer la précision du contrôle, de réduire les pertes et d’optimiser la production.

Cependant, après le déploiement du modèle, un problème majeur est apparu : toutes les vis sont classifiées comme conformes, remettant en question l’efficacité du modèle et compromettant la chaîne de production.

### 

### **2. Objectifs du Projet**

L’objectif principal du projet est d’améliorer la précision de détection des défauts sur les vis en :

* Corrigeant les erreurs du modèle existant.
* Ajustant les seuils de classification pour améliorer la différenciation entre vis conformes et non conformes.
* Optimisant l’apprentissage du modèle en tenant compte des biais dans les données.
* Assurant une meilleure traçabilité et une documentation complète du processus.

### 

### **3. Problématique**

Le modèle actuel est inefficace, ne parvenant pas à identifier les vis défectueuses. Plusieurs causes potentielles sont envisagées :

* Un jeu de données déséquilibré favorisant l’apprentissage sur les vis conformes.
* Des seuils de classification mal ajustés.
* Un sur-apprentissage dû à un manque de diversité dans les images utilisées.

Une analyse approfondie et une série d’expérimentations seront nécessaires pour proposer des solutions adaptées.

**Remarque importante :** Lors de la phase de test, deux jours ont été consacrés à l’expérimentation du modèle. Cependant, le modèle n’a pas été suffisamment entraîné sur les dossiers de validation, d'entraînement et de test. Cela a fortement limité sa capacité de généralisation. En conséquence, bien que les premières métriques semblaient encourageantes, le modèle n’a pas atteint un niveau de fiabilité satisfaisant (100%) en situation réelle.

Ce manque d'entraînement approfondi explique en partie pourquoi certaines vis défectueuses n’ont pas été correctement identifiées.

### **4. Ressources Fournies**

* **Référentiel RNCP 36146** : pour garantir la conformité du travail aux standards professionnels.
* **Un dépôt GitHub contenant :** <https://github.com/HorHakim/Screw>
  + Le code d’entraînement du modèle tel qu’il est utilisé aujourd’hui.
  + [Une application Streamlit de démonstration](https://okvisfactory.streamlit.app/) : cette application déployée en production permet de soumettre une image de vis et d’obtenir un diagnostic du modèle.

### 

### **5. Modifications et Justifications**

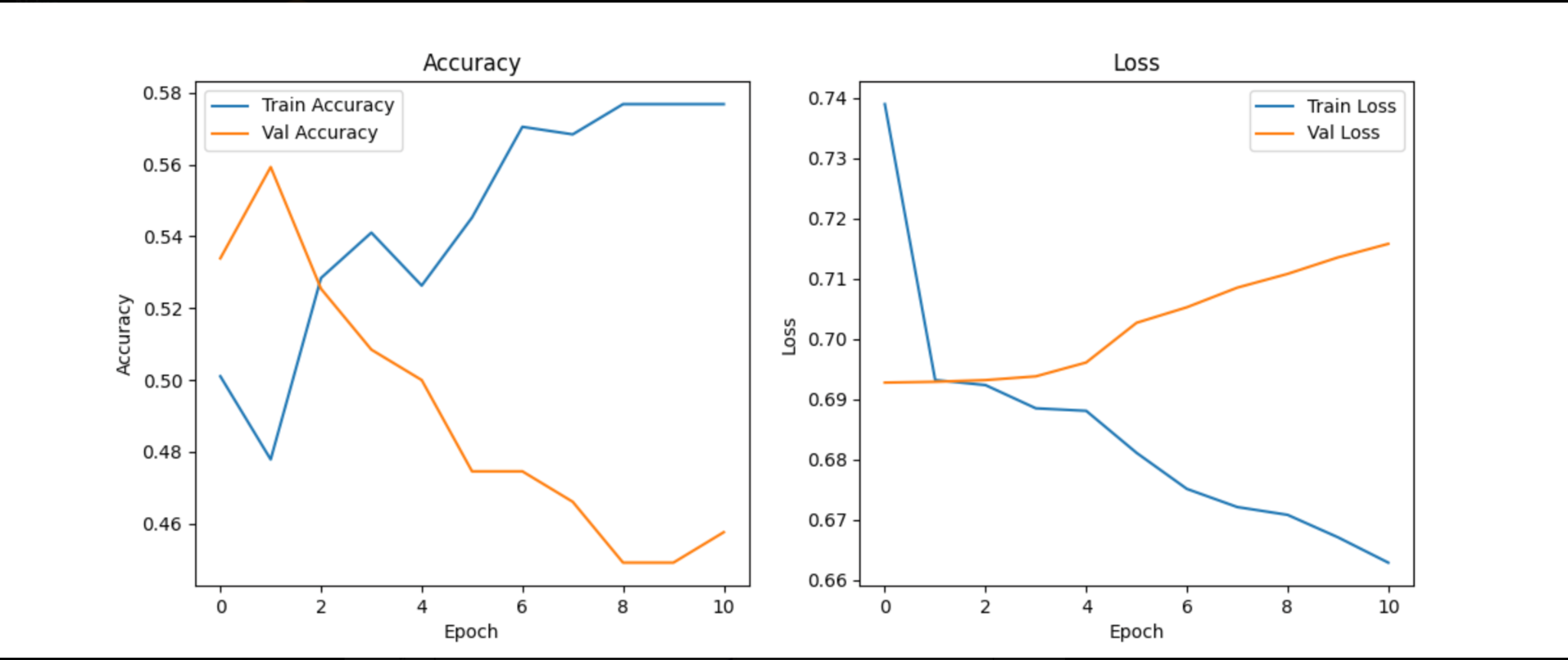
#### **a) Comparaison des Modifications**

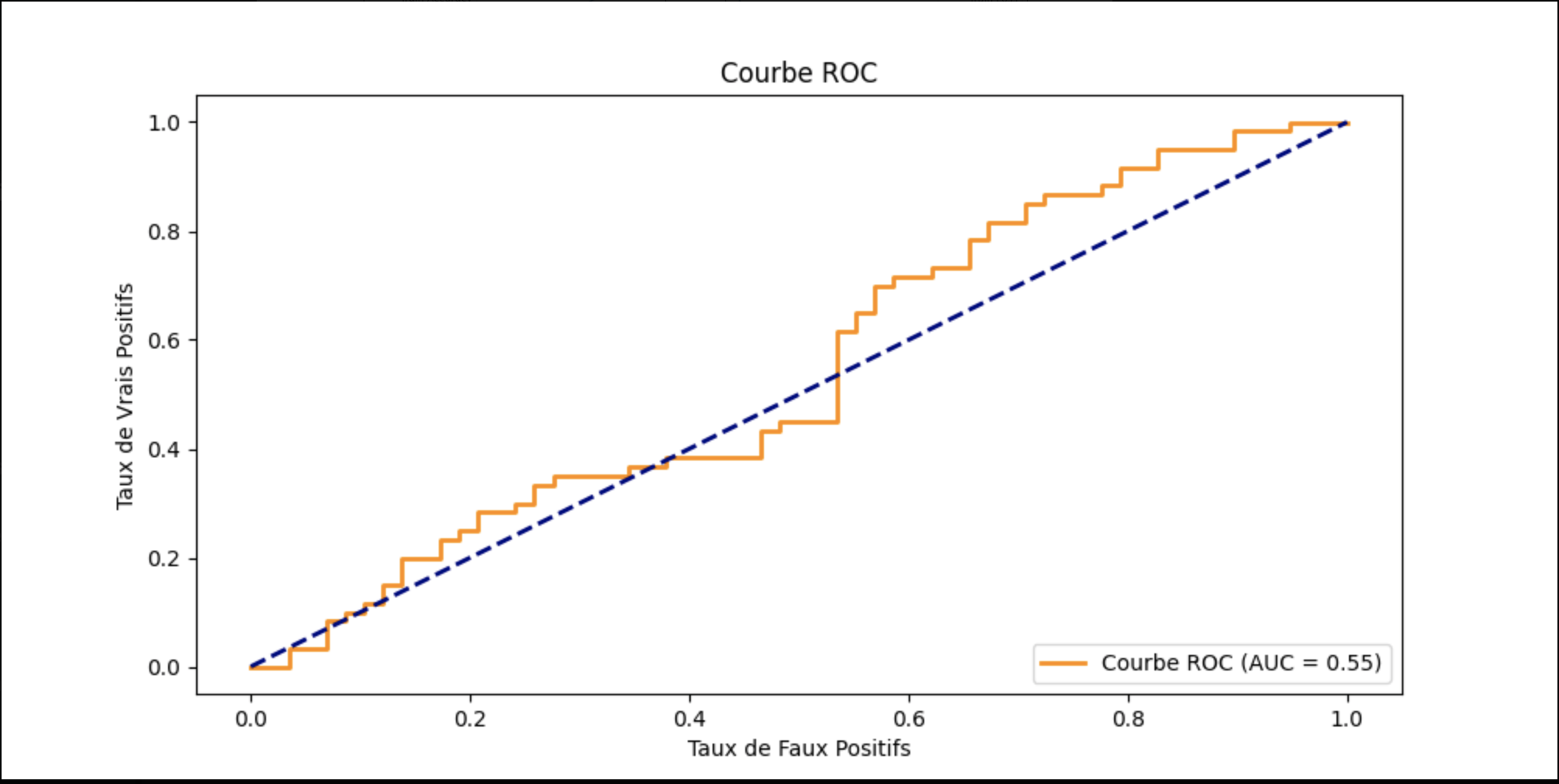
* **Categorical Cross Entropy vs Binary Cross Entropy** :
  + Nous avons choisi **Binary Cross Entropy (BCE)** comme fonction de perte car le problème est une classification binaire (vis conforme ou défectueuse).
  + **Categorical Cross Entropy** est utilisée pour des problèmes multi-classes, ce qui n’est pas pertinent ici.
* **Séparation du Train et de la Validation** :
  + Séparation explicite du dataset en **80% pour l'entraînement et 20% pour la validation** pour éviter le sur-ajustement.
  + Mise en place d’une **shuffle des données** pour garantir un bon échantillonnage et éviter un biais de distribution.
* **Réduction des Epochs et Early Stopping** :
  + Pour optimiser le temps d'entraînement et éviter un sur-apprentissage, nous avons **réduit le nombre d'epochs et** ajouté **l'early stopping**.[[2]](#footnote-1)
  + **Early stopping** surveille la perte sur l’ensemble de validation et stoppe l'entraînement lorsqu’il n’y a plus d'amélioration après un certain nombre d'epochs.
* **Ajout de Couches CNN** :
  + Augmentation du nombre de couches convolutionnelles (CNN) pour améliorer l’extraction des caractéristiques des vis.
  + Ajout de Batch Normalization et Dropout pour stabiliser l’apprentissage et éviter le sur-apprentissage.
* **Augmentation de Données (Data Augmentation)** :
  + Ajout d’une augmentation de données pour générer plus d'exemples et améliorer la généralisation du modèle.
  + Techniques utilisées : rotation, zoom, flip horizontal, modification de contraste et bruit gaussien.

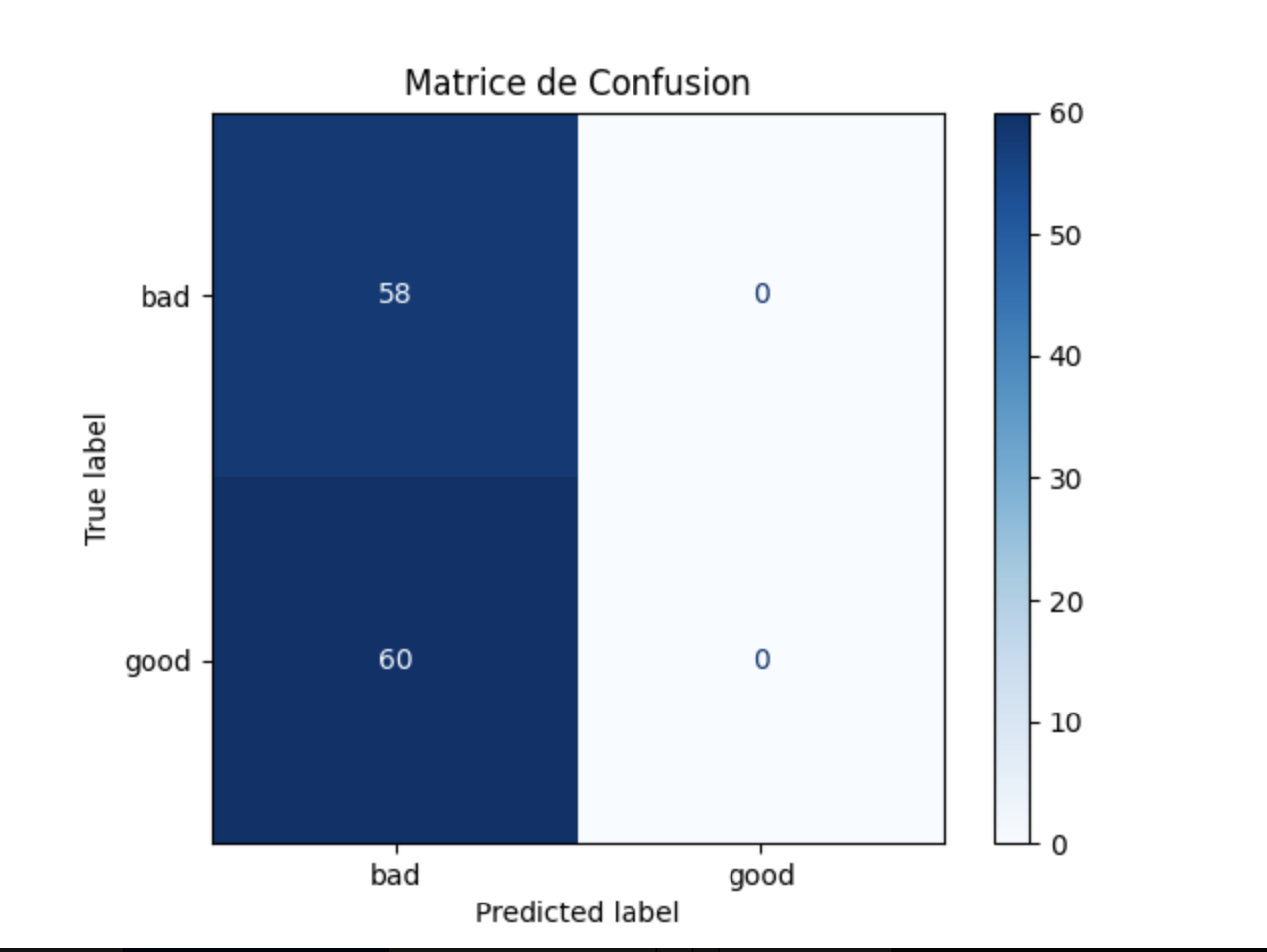
#### **b) Analyse des Performances du Modèle**

L’évolution de l’accuracy et de la loss est illustrée dans le graphe ci-dessous :

* **Courbe ROC** :
  + La courbe ROC ci-dessous montre un **AUC de 0.55**, ce qui indique une faible capacité de discrimination du modèle.
  + Un score proche de **0.5** signifie que le modèle n'est pas meilleur qu’un choix aléatoire.
* **Matrice de Confusion** :
  + La matrice de confusion montre que le modèle classe **toutes les vis comme mauvaises**.
  + Aucun vrai positif pour les vis conformes, ce qui signifie une **mauvaise calibration du modèle**.
  + Cela confirme un **problème de généralisation**, nécessitant des ajustements sur le seuil de classification.
* **Accuracy :**
  + La **train accuracy** (bleu) augmente progressivement, ce qui montre que le modèle apprend bien sur l’ensemble d'entraînement.
  + La **validation accuracy** (orange) baisse après quelques epochs, ce qui indique un **problème de sur-apprentissage (overfitting)**. Le modèle devient trop spécialisé sur les données d’entraînement et perd sa capacité de généralisation.
  + Un **écart trop important** entre la précision d'entraînement et celle de validation est un signe que le modèle mémorise les données au lieu d’apprendre des patterns robustes.
* **Loss :**
  + La **train loss** (bleu) diminue de manière continue, ce qui est attendu lorsqu’un modèle apprend bien.[[3]](#footnote-2)
  + La **validation loss** (orange) augmente après quelques epochs, ce qui indique une **détériorisation des performances en généralisation**.
  + Ce comportement est typique d’un modèle qui commence à **sur-ajuster aux données d'entraînement**, d’où l’importance d’un **early stopping et d’une régularisation**.



[[4]](#footnote-3)



**Pourquoi ces modifications ?**

* **Séparation Train/Validation** : évite le sur-apprentissage et améliore la robustesse du modèle.
* **Early Stopping** : évite un entraînement trop long et inutiles.
* **Plus de couches CNN** : améliore la reconnaissance des motifs et augmente la performance.
* **Augmentation des données** : permet d'entraîner un modèle plus robuste même avec un dataset limité.[[5]](#footnote-4)

### 

### 

### 

### 

### 

### **6. Livrables**

* **Code source optimisé** : avec documentation et annotations.
* **Rapport d’analyse** : détaillant les problèmes identifiés et les solutions mises en place.
* **Rapport de performance** : incluant les résultats obtenus après correction, l’analyse de la courbe ROC et de la matrice de confusion.
* **Présentation pour les décideurs** : mettant en avant les résultats et recommandations.

### 

### **7. Critères de Réussite**

* Amélioration notable du taux de détection des vis défectueuses.
* Réduction des erreurs de classification et augmentation de la précision.
* Documentation complète et claire, conforme aux standards du RNCP 36146.
* Mise en place d’un pipeline reproductible et adaptable pour de futures améliorations.
* Utilisation de la courbe ROC pour justifier le choix du seuil de classification optimal.
* Vérification des performances via la matrice de confusion et l’analyse des faux positifs/négatifs.[[6]](#footnote-5)
* Optimisation de la fonction de perte avec la Binary Cross Entropy et de la sortie du modèle avec la Sigmoïde.[[7]](#footnote-6)
* Augmentation des données et ajout de nouvelles couches CNN pour améliorer l'extraction de caractéristiques.
* Mise en place de l’early stopping pour éviter un sur-ajustement.

### 

### 

### **8. Conclusion**

Ce projet vise à garantir un contrôle qualité automatisé fiable et performant en corrigeant les failles du modèle initial. En combinant une analyse approfondie, des ajustements techniques et une méthodologie rigoureuse, nous proposons une [[8]](#footnote-7)solution robuste et conforme aux exigences du secteur industriel.[[9]](#footnote-8)

1. [↑](#footnote-ref-0)
2. [↑](#footnote-ref-1)
3. [↑](#footnote-ref-2)
4. [↑](#footnote-ref-3)
5. [↑](#footnote-ref-4)
6. [↑](#footnote-ref-5)
7. [↑](#footnote-ref-6)
8. [↑](#footnote-ref-7)
9. [↑](#footnote-ref-8)