AI for Pig Disease Classification

Ai in society

NOVEMBER 27, 2024

LUCAS WEIJS

Contents

Abstract

Pig disease classification is a pivotal challenge in the agricultural and veterinary sectors, impacting animal health, farm productivity, and food security. Leveraging Artificial Intelligence (AI), this project aims to develop a robust classification system to detect and diagnose pig diseases with high accuracy. The research integrates advanced machine learning techniques, comprehensive data preprocessing, and model optimization to address the complexities of real-world applications.

The study emphasizes sustainability, ethical considerations, and societal impacts while adhering to data legislation. Challenges such as data quality, imbalance, and model interpretability are critically analyzed, with proposed solutions to mitigate these issues. Additionally, interactive visualizations are employed to effectively communicate findings to stakeholders, including farmers and industry professionals.

The results demonstrate the potential of AI to revolutionize disease management in agriculture, contributing to sustainable practices and enhanced animal welfare. This research sets the stage for future advancements in AI-driven veterinary diagnostics, aligning with broader goals of intelligent and responsible farming innovations.

1. Sustainable Context and Societal Impact

The application of AI for pig disease classification tackles several pressing sustainability and societal challenges, particularly in agriculture and veterinary medicine. By using AI to improve the early detection and accuracy of disease identification, this project offers a transformative solution that not only enhances animal health but also aligns with broader global goals for sustainability and ethical practices in farming.

1. **Improved Early Detection and Targeted Interventions**
   AI-driven disease classification enables farmers and veterinarians to identify health issues in pigs at an earlier stage, significantly improving the response time for treatment. This early detection allows for more precise, targeted interventions rather than broad-spectrum treatments, such as antibiotics. As a result, the need for indiscriminate medication is reduced.

2. **Promotion of Animal Welfare**
   Timely and accurate disease detection ensures that animals receive appropriate care before health issues progress to more serious stages. This not only contributes to better outcomes for the animals but also fosters overall animal welfare. The use of AI provides a more objective and consistent approach to monitoring animal health, reducing the potential for human error in assessments and ensuring that animals are treated promptly, leading to better quality of life for livestock.

3. **Optimized Resource Utilization and Sustainable Food Production**
   The increased efficiency in farming practices resulting from improved disease management also leads to optimized resource utilization. Healthier livestock produces more meat, milk, or other products while consuming fewer inputs. This contributes to the broader goals of sustainable food production, where systems are designed to maximize output while minimizing waste and the environmental footprint. By ensuring that animals are kept healthy and disease-free, AI can play a key role in improving overall food security and reducing the environmental impact of livestock farming.

4. **Ethical Considerations and Responsible Use of AI**
   The project emphasizes the ethical responsibility in applying AI, ensuring that the technology is used transparently and in compliance with data protection and legislation. The data collected for training the AI models is handled responsibly, with the privacy and rights of individuals considered at every step. This ethical part not only builds trust in AI but also promotes responsible adoption across farming and veterinary practices. By empowering farmers and veterinarians with data-driven, reliable insights, the project reduces reliance on manual assessments, which are often time-consuming and subject to error. This leads to more informed decision-making, improving both the efficiency and fairness of veterinary practices.

5. **Broader Societal Impact**
   Beyond the direct benefits to individual farms, the societal impact of this project is significant. By minimizing the spread of diseases within herds, the AI system plays a critical role in safeguarding public health. Livestock diseases that go undetected or unchecked can have far-reaching consequences, including potential transmissions to humans. Early identification and intervention reduce the likelihood of outbreaks that could affect not just the farm but the broader community. This contribution to public health underscores the importance of leveraging AI to address challenges that have a direct and indirect impact on society.

In conclusion, the use of AI for pig disease classification serves as a powerful tool to address the challenges of sustainability, animal welfare, and public health. It helps farmers and veterinarians make more informed, efficient, and ethical decisions, contributing to the broader goals of sustainable agriculture and responsible use of technology. By improving early disease detection and minimizing the spread of infections, this project supports the ongoing effort to create more resilient, efficient, and ethical farming systems.

2. Critical Analysis of the AI Project

Challenges and Solutions in the AI Project

Throughout the course of this project, several challenges emerged, primarily related to the limitations in data and the generalizability of the AI model. These issues were identified early in the project and addressed with targeted solutions, ultimately improving the model's performance and ensuring its robustness.

1. **Challenges Encountered**
   The AI project faced several significant challenges:
   - **Data Scarcity:** At the start, there was a limited amount of high-quality, labeled data available for pig disease classification. This made training an accurate model difficult.
   - **Model Generalizability:** The initial models struggled to perform well on unseen data, indicating overfitting and limited robustness.
2. **Implemented Solutions**
   - **Data Augmentation:** To address the issue of limited data, I applied techniques such as rotation, flipping, zooming, and brightness adjustments. These techniques artificially expanded the dataset, helping the model become more adaptable to new and varied images.
   - **Transfer Learning:** I employed transfer learning by using the pre-trained VGG16 model. This allowed the model to leverage existing knowledge from large-scale image datasets. By fine-tuning the model for pig disease classification, I effectively addressed the generalizability issue and boosted performance.

Roadmap of Challenges and Solutions

Weeks 1–3: Initial Challenges with Data and Models

- **Focus:** Collected initial datasets and built a custom neural network. However, the model's performance was poor due to data scarcity and limited diversity.
- **Outcome:** Identified the need for more data and better model strategies to improve results.

Weeks 4–5: Exploring and Implementing Solutions

- **Focus:**
    - Introduced data augmentation techniques to increase dataset size and variability.
    - Transitioned to the VGG16 model using transfer learning to overcome the limitations of the initial custom model.
- **Outcome:** Improved dataset quality and enhanced model performance on unseen data.

Weeks 6–8: Model Optimization and Testing

- **Focus:**
    - Fine-tuned the VGG16 model for pig disease classification.
    - Applied advanced techniques like hyperparameter tuning and dropout to prevent overfitting and boost accuracy.
- **Outcome:** Achieved a robust model capable of handling the complexities of the task, with better generalizability across diverse test cases.

This structured approach allowed me to systematically identify and resolve the challenges, ensuring the project's success in building an accurate and generalizable AI model. By addressing data limitations through augmentation and leveraging transfer learning, I was able to overcome initial setbacks and deliver a model with strong performance and real-world applicability.

3. Data Collection and Preparation

Data collection forms the backbone of the project. This chapter details the sourcing of high-quality data, its preprocessing, and adjustments made to enhance usability for AI models. Processes such as cleaning, normalization, and annotation are explained.

Data collection formed the foundation of this project, emphasizing the importance of high-quality and balanced datasets for effective AI model performance. To ensure the dataset's usability, an extensive manual review of images was conducted to remove poor-quality samples that could hinder model training. Multiple datasets were considered during this phase; however, the chosen dataset was selected based on its prior use in another AI project, which had already demonstrated a baseline accuracy of 87%. This provided confidence in its reliability and relevance to the task.

The scarcity of openly available data on pig diseases posed a significant challenge. To address this, the dataset was expanded to include diverse visual diseases rather than limiting the classification to binary categories like "healthy" and "ill." This approach not only enhanced the dataset’s depth but also better reflected real-world complexities. Preprocessing steps such as cleaning, normalization, and annotation were rigorously applied, while tools like ImageDataGenerator were employed to automate data preparation. The dataset was ultimately split into training (80%) and validation (20%) subsets to ensure balanced and effective learning.

Data Collection and Refinement Roadmap

1. **Initial Challenges**
   At the start, the project faced significant issues with incorrect data and an insufficient number of samples. This limited the model's ability to generalize and impacted early experiments with AI training.

2. **Exploration and Dataset Improvement**
   To address these challenges, I began exploring multiple datasets to find those that could provide better quality and quantity. After thorough evaluation, I discovered a dataset that had been used in a similar AI project, achieving a baseline accuracy of 87%. This dataset became the foundation for the later stages of the project.

3. **Cleaning and Quality Control**
   With the new dataset in hand, I conducted a manual review to identify and remove poor-quality or irrelevant samples. This ensured the data was clean, accurate, and suitable for training.

4. **Dataset Expansion**
   Recognizing gaps in the dataset, I expanded it by incorporating images representing a broader range of pig diseases. This added diversity and moved beyond basic binary

classifications like "healthy" and "ill," making the dataset more realistic and comprehensive.

5. **Preprocessing**
   I applied preprocessing techniques such as normalization, resizing, and augmentation to prepare the data for training. Tools like ImageDataGenerator were used to streamline and automate these processes.

6. **Annotation**
   The dataset was annotated to create precise labels, ensuring the AI model could learn effectively from the available data.

7. **Splitting and Iteration**
   I divided the data into 80% training and 20% validation subsets for balanced learning. During training, I revisited the dataset, identified biases or errors, and refined the data as needed to maintain accuracy and relevance.

4. Model Training and Testing

This section covers the selection of AI models, training methodologies, and testing strategies. Evaluation metrics such as accuracy, precision, and recall are used to assess model performance. Hyperparameter tuning and cross-validation techniques are also discussed.

The VGG16 model architecture was selected due to its well-documented success in image classification tasks. Transfer learning techniques were employed, leveraging VGG16's pre-trained layers while adding custom layers tailored to the pig disease classification task. Training started with an initial accuracy of 15.4% in the first epoch, gradually improving as the model learned to distinguish between diverse disease categories.

This project explored various models, including custom Convolutional Neural Networks (CNNs) and fully connected neural networks. While these models showed promise, they failed to match the performance of VGG16, which consistently achieved the highest accuracy. The inclusion of multiple disease categories instead of a simple binary classification played a critical role in enhancing the model's predictive power.

Advanced training methodologies were applied, including hyperparameter tuning and cross-validation, to optimize model performance. Regularization techniques such as dropout were also employed to prevent overfitting. The results reaffirmed that VGG16, combined with a thoughtfully curated and diverse dataset, was the most effective solution for this task.

Model Selection, Training, and Evaluation Roadmap

1. **Initial Attempt with a Custom Neural Network**
   I began by designing and training a custom neural network tailored for pig disease classification. While this approach provided valuable learning experience, a review with my teachers revealed that the model's accuracy was insufficient for reliable predictions. Its limitations in handling complex image recognition tasks highlighted the need for a more robust solution.

2. **Exploring Alternatives**
   Based on feedback and further research, I explored established models like VGG16, YOLOv5, and Support Vector Machines (SVM). These models are widely recognized for their strong performance in image recognition tasks and offered a solid foundation for the project.

3. **Switching to VGG16**
   After careful consideration, I selected the VGG16 model due to its well-documented success in image classification. Leveraging its pre-trained layers and customizing the architecture for pig disease classification provided a significant performance boost compared to the custom neural network.

4. **Experimenting with YOLOv5**
   YOLOv5 was considered for its real-time object detection capabilities, which could enhance the model's ability to identify specific features related to pig diseases. This added another dimension to the evaluation process.

5. **Incorporating Support Vector Machines**
   SVMs were explored as a complementary approach for simpler image recognition tasks. While less suited for the complexity of pig disease classification, they provided valuable insights into feature separation and classification.

6. **Final Decision**
   VGG16 emerged as the most effective solution, balancing high accuracy with scalability for the task. The switch from a custom neural network to a more established architecture ensured reliable results and better alignment with the project’s goals.

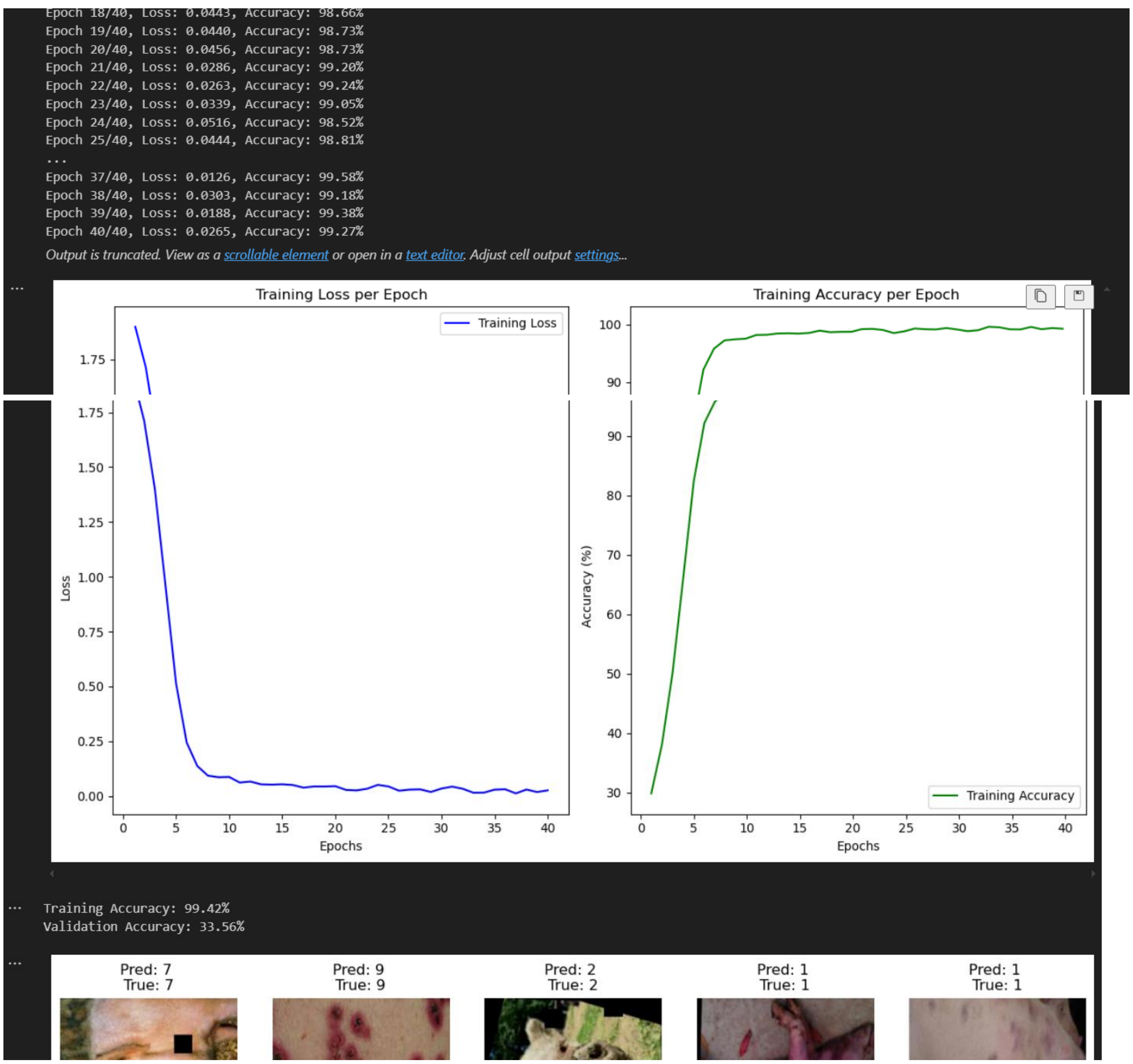
5. Visualization and Communication

Creating engaging and informative visualizations is key to communicating findings. This chapter presents examples of data storytelling tailored to stakeholders, including farmers and researchers.

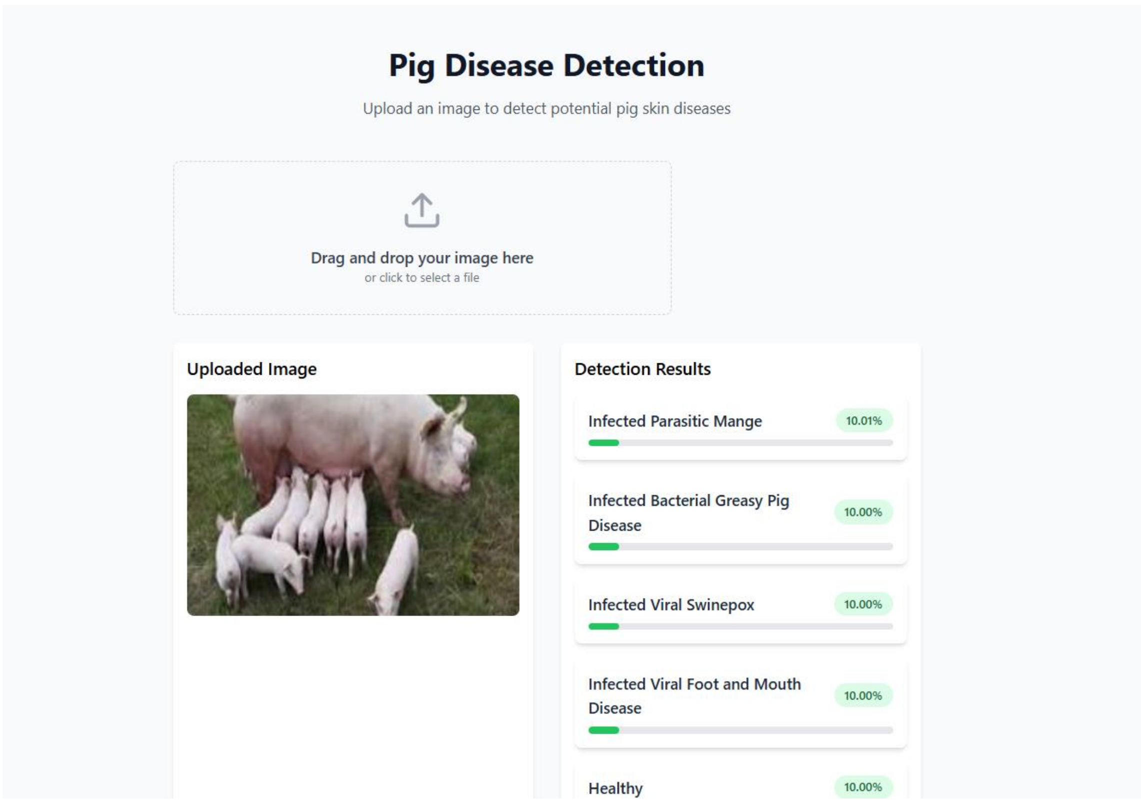
Engaging Visualizations and AI Integration

1. **I started to visualize the training:**

   this showed me early that the model I was using was not correct and helped me learn about epochs and overfitting.



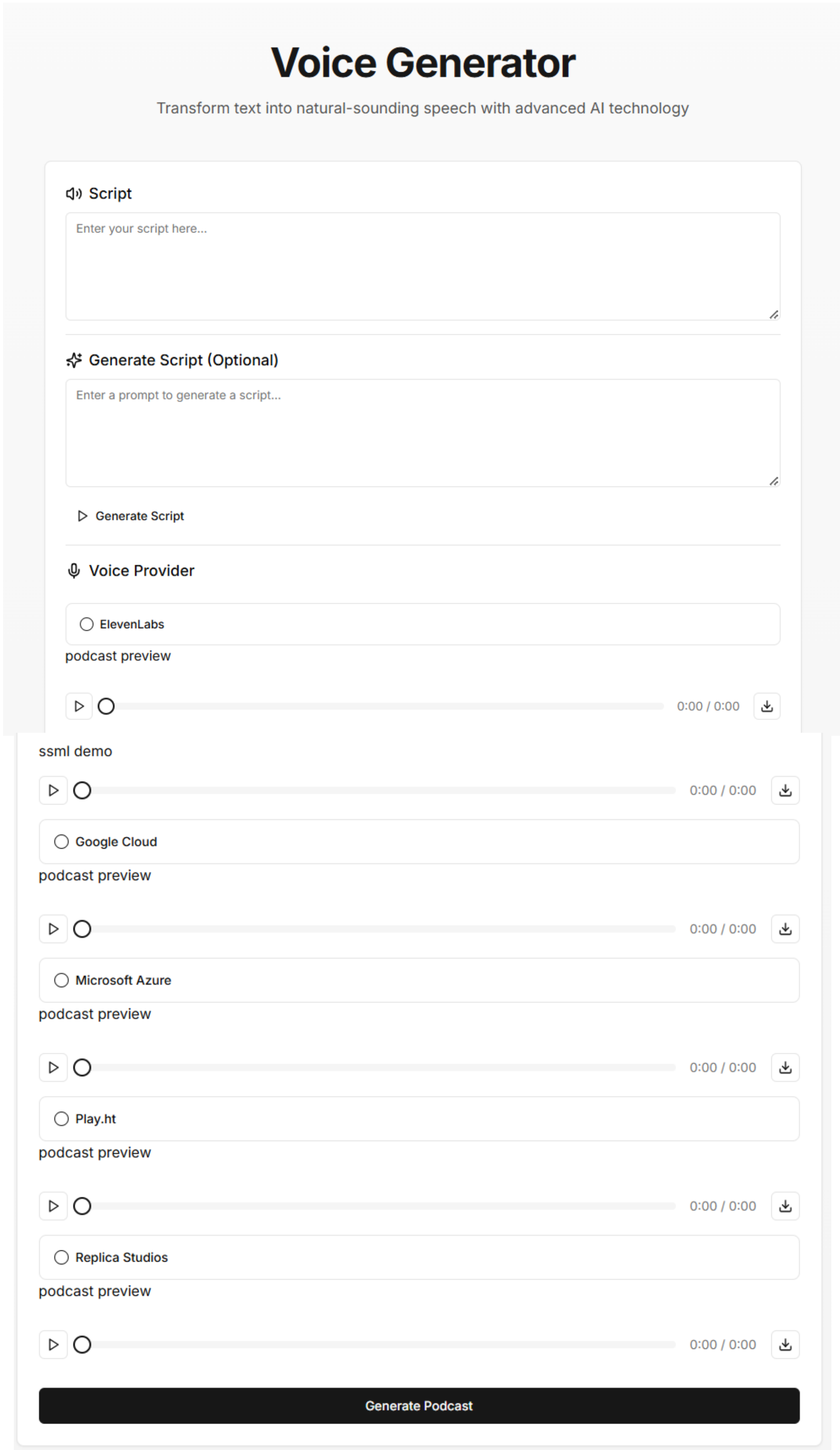
2. **AI Integration with a Basic Website**
   To make the AI algorithm more accessible, I developed a basic website that allows users to upload images and receive classification results directly from the AI model. The website features a simple interface tailored to its target users, such as farmers and researchers, ensuring ease of use while delivering accurate predictions.



3. **Visualization and Data Storytelling**
   The website includes visualizations to make the findings more engaging and informative. Charts, graphs, and annotated image outputs from the AI model help stakeholders, such as farmers and researchers, better understand the predictions and insights generated by the algorithm. These visualizations bridge the gap between raw data and actionable knowledge.

4. **MEDIA 52**

   Additionally for the project we made a website to visualize the voices so we can do better research and test different competitors and also try custom prompts with or without ssml(the markup language for voices)


Voice Generator
Transform text into natural-sounding speech with advanced AI technology
Script
Enter your script here...
Generate Script (Optional)
Enter a prompt to generate a script...
Generate Script
Voice Provider
ElevenLabs
podcast preview
0:00 / 0:00
ssml demo
0:00 / 0:00
Google Cloud
podcast preview
0:00 / 0:00
Microsoft Azure
podcast preview
0:00 / 0:00
Play.ht
podcast preview
0:00 / 0:00
Replica Studios
podcast preview
0:00 / 0:00
Generate Podcast

6. Project Reporting and Reflection

Throughout the project, I made a effort to document and present my progress, ensuring that both my understanding and outcomes were clearly communicated to my team and stakeholders. Several tools were used to track the project's evolution and ensure a clear reflection on the work done.

1. **Python Notebook for Progress Tracking**
   To document my experiments and track the progress of the AI model, I created a detailed Python notebook. This notebook served as an interactive report, showcasing key stages of the project, such as data preprocessing, model training, and evaluation. It provided a step-by-step guide to my thought process and the adjustments made during development, offering full transparency of the workflow.

2. **Research Paper for In-Depth Analysis**
   To better illustrate my understanding and the methodologies applied throughout the project, I wrote a research paper. The paper highlighted the challenges faced, the solutions implemented (such as data augmentation and transfer learning), and the results of various model evaluations. It acted as a comprehensive summary of the project, providing insights into both the technical and conceptual aspects.

3. **Project Documentation for Team Clarity**
   Additionally, I created several documents to ensure the clarity of the work for my group. These documents summarized key research findings, AI model choices, and relevant project decisions. By organizing and structuring this information, I ensured that all team members had access to the necessary details to understand the project's direction and progress. These documents acted as a reference point, enabling effective collaboration and knowledge sharing within the team.

4. **Using Jira in Scrum Methodology**
   To manage tasks and maintain organized project flow, we implemented Scrum methodology using Jira. This allowed us to break down the project into manageable sprints, set specific deliverables, and track progress effectively. By using Jira, we were able to collaborate efficiently, stay on top of deadlines, and ensure that everyone on the team was aligned with the project's goals. It also facilitated transparency, as tasks, bugs, and milestones were clearly outlined and regularly updated.

By combining technical documentation, research-based analysis, and clear communication with the team, I was able to maintain a comprehensive record of the project and reflect on its outcomes in a structured and accessible way.

7. Entrepreneurial Insights and Professional Development

The project demonstrates an entrepreneurial mindset by addressing real-world challenges in animal health. This chapter reflects on personal growth and aligns the project with future professional ambitions.

8. Conclusion and Future Work

This paper concludes by summarizing the outcomes of the AI-based pig disease classification project. Future work aims to refine model accuracy and expand its applicability across different contexts. The long-term vision includes integrating AI solutions into broader veterinary practices.