CHAPTER 6

Conclusions, Discussion and Future Work

6.1 Chapter Summary

In this thesis, we describe our research in four chapters, each addressing a specific set of research questions. Our primary focus has been the exploration and application of deep learning techniques in various fields, ranging from the intricacies of agriculture and the complexities of biomedical research to the innovative intersection of image processing with Natural Language Processing. Through this diverse exploration, we contributed to the body of knowledge in these individual domains as well as demonstrated the versatile applicability and transformative potential of deep learning technologies.

In this chapter, we focus on weaving together the diverse strands of our research, providing a synthesis of the main findings that have emerged from each chapter. We reflect on the breakthroughs and insights gained, particularly focusing on how our approaches have pushed the boundaries of what is possible with deep learning strategies in various applications. In doing so, we also acknowledge the limitations and challenges encountered, offering a critical evaluation of our methodologies and their implications.

Furthermore, this chapter serves as a springboard for future research directions. It outlines the potential pathways and opportunities that our work has opened up, suggesting how future researchers can build upon our findings to explore new frontiers. The intent is to provide a road map that guides subsequent explorations, encouraging the continued evolution and innovation in the fascinating fields of deep learning, image analysis, and beyond.

6.2 Main Findings

In the following paragraphs, we will address our research questions, providing detailed explanations and insights for each.

6.2.1 RQ1

In Chapter 2, we focused on feature extraction task through CNNs and dedicated our efforts addressing the first set of research questions (RQ1):

RQ1: *Is it possible to accurately classify different corn-seed varieties by analyzing single-instance seed images using Deep Learning, focusing on specific image-derived features?*

SubQ1: *In the context of successful classification, which specific features extracted by Deep Learning algorithms contribute most to determining the accuracy of these classifications?*

These research questions emerged from the critical need in agricultural biotechnology for precise and efficient methods to classify corn seed varieties, a task that has significant implications for crop yield and seed purity.

To address this challenge, we developed and implemented a deep CNN as a feature extractor. This approach represents a significant shift from the traditional methods that utilize handcrafted features, relying predominantly on basic computer vision techniques and manual inspection. Our CNN model is specifically engineered to automatically extract complex features from corn seed images, identifying nuances and details that might be missed by human inspection or simpler algorithms. We further augmented the capabilities of CNN by integrating it with various sophisticated classification algorithms, such as artificial neural networks, support vector machines, and k-nearest neighbors. This combination of deep learning for feature extraction with advanced classification techniques was pivotal in achieving our goal. The CNN-ANN classifier, in particular, stood out as a good model, delivering exceptional performance metrics. It demonstrated remarkable accuracy and efficiency, classifying a vast number of seed varieties swiftly and with a high degree of precision.

Our findings in Chapter 2 affirmed RQ1 and also set a new benchmark in the field of seed classification technology. The CNN-based model we developed proved capable of handling the complexity of corn seed variety classification with unprecedented accuracy. This success paves the way for similar deep learning-based approaches to be applied to other seed species, potentially revolutionizing seed classification and agricultural practices broader. The adaptability and scalability of our model meant that it can be tailored to a wide range of agricultural Biotech applications, making it a valuable tool for stakeholders such as seed companies and horticultural producers.

In Chapter 2 we successfully addressed RQ1 by demonstrating that CNNs, when combined with advanced classification algorithms, can effectively classify corn seed varieties. The high accuracy and efficiency of the models in feature extraction and classification have significant implications for agricultural practices, offering a more reliable, scalable, and precise method compared to traditional approaches.

6.2.2 RQ2

In Chapter 3, we delved into the second set of research questions (RQ2) focusing on image classification through CNNs. The application of this classification task is essential in advancing horticulture practices in agriculture technology:

RQ2: *Can Deep Learning techniques effectively determine the ripeness stage of mulberries by analyzing single-instance images of these fruits?*

SubQ2: *Assuming successful ripeness stage classification, what would be the consequent impact on the efficiency and effectiveness of post-harvest processing in mulberries?*

These questions arose from the significant challenges in the horticultural sector and Agri-food industry, particularly the sorting and classification of fruits based on ripeness - a task traditionally dependent on manual labor and subjective evaluation. In this chapter, our research focused on utilizing advanced CNN architectures, namely AlexNet and ResNet-18, for the precise determination of mulberry ripeness stages. Mulberries are selected for their economic significance and their clearly defined ripening stages, which present an interesting classification challenge due to subtle variations in color and texture.

Our findings are revolutionary. The AlexNet and ResNet-18 models exhibited outstanding accuracy and efficiency in categorizing mulberries into various stages of ripeness. These models adeptly recognized the subtle shifts in appearance that occur as the fruit ripens, a task that is challenging for human sorters and traditional methods. This automated approach using CNNs significantly surpasses human accuracy, reducing errors and the consequent waste from misclassified fruits. Moreover, the success of these models in discerning mulberry ripeness suggests their applicability to other fruits and horticultural products. Their proficiency in learning and adapting to diverse textures, colors, and shapes renders them invaluable for a broad spectrum of agricultural sorting and classification tasks.

Addressing RQ2, our research not only confirmed the effectiveness of CNNs in classifying mulberry ripeness stages but also underscored the far-reaching impact of this technology in enhancing post-harvest processing. By minimizing waste and maximizing sorting efficiency, our approach contributes markedly to more sustainable practices in the field of post-harvest technology in the agriculture industry. The scalability of this technology to other fruits and crops heralds a potential revolution in the agricultural sector, promising a more objective, accurate, and efficient method for classifying ripeness.

6.2.3 RQ3

In Chapter 4, we addressed the third set of research questions (RQ3) focusing on image segmentation task through CNNs:

RQ3: *How effective and swift can Deep Learning be in segmenting images of zebrafish larvae obtained from microscope setups, particularly in distinguishing larvae from their surrounding environment?*

SubQ3: *If such segmentation is achieved effectively, what are the potential impacts on High-Throughput Screening (HTS) pipelines and multi-dimensional imaging processes in research settings?*

These questions are important given the pivotal role of precise image segmentation in biomedical research. This is particularly true for microscopic images of model organisms like zebrafish, widely used in biomedical studies.

In our approach, we made a significant advancement in biomedical image processing. We integrated the Transformer model, acclaimed for its proficiency in capturing global dependencies, with the U-Net architecture, which is renowned in the field of biomedical image segmentation. This innovative combination aimed to address the shortcomings of conventional segmentation methods that typically falter with the complex and varied nature of biomedical imagery.

The outcomes of this chapter are notably promising. Our hybrid Transformer-U-Net model exhibited exceptional accuracy and precision in segmenting images of zebrafish larvae. Such precision is indispensable for the next phase of generating 3D reconstructions from 2D images, a process that demands meticulous attention to detail for accuracy. Our results suggest that this methodology could markedly enhance the quality of biomedical analysis, yielding clearer and more intricate images that could lead to deeper insights and understanding in biomedical research.

Furthermore, the efficacy of transfer learning in segmenting zebrafish larvae images indicates its potential for broader application in biomedical imaging. This technique could be tailored for other model organisms or various types of biomedical images, thereby potentially revolutionizing the accuracy and efficiency of image analyses across diverse biomedical research fields.

6.2.4 RQ4

In Chapter 5, we addressed the fourth research question set (RQ4), focusing on the integration of image processing with natural language processing (NLP) through CNNs:

RQ4: *Can the integration of Deep Learning (focused on image analysis) and NLP lead to significant improvements in image captioning techniques?*

SubQ4: *If so, what are the expected advancements in terms of performance and accuracy that could be realized across various domains, such as medical imaging, surveillance, and digital media?*

In our previous chapters, we primarily concentrated on processing at the pixel level, focusing on the visual aspects of images. However, when considering the content of images in terms of the concepts they portray, the semantics embedded in these concepts become pivotal. Image captions are instrumental in providing this semantic context. In the realm of image captioning, the key challenge lies in bridging the gap between visual recognition and language processing.

Chapter 5 was dedicated to developing and evaluating a groundbreaking framework for image captioning, transcending the conventional bounds of CNN-based models. We acknowledged that while CNNs excel in object detection, they cannot often accurately interpret the context and relationships within images. To overcome this limitation, we introduced a novel model that synergistically combines advanced neural network architectures, harnessing their strengths in both visual perception and linguistic understanding. The findings from this chapter showcased an exceptional proficiency in not only detecting objects within images but also in accurately contextualizing and relating them. This represents a substantial leap forward from the traditional CNN-based methods, which generally falter in these more nuanced aspects of image captioning.

The triumph of our model in Chapter 5 heralds a new era for image captioning applications. It underscores the potential for creating more sophisticated, context-aware captioning systems. Such systems could prove invaluable across various domains, ranging from improving accessibility for the visually impaired to offering enhanced content analysis on digital media platforms. The advancements in performance and accuracy, as explored in SubQ4, could significantly benefit fields like medical imaging and surveillance, where precise and contextually rich interpretations of visual data are crucial.

6.3 Future work

6.3.1 Agronomy applications (RQ1 and RQ2)

In the field of postharvest technology in the agriculture industry, particularly in the realms of identifying seed varieties and determining the ripeness level of horticultural products, a variety of technologies have been employed. These technologies range from traditional manual methods (Chakraborty et al., 2022), to more advanced technological solutions like using deep learning models in both seed classification (Han et al., 2024), (Zhang et al., 2024) and fruit ripeness detection (Ge et al., 2024) domains. Currently, there is an increasing focus on incorporating machine learning and image processing techniques to improve accuracy and efficiency. Despite these advancements, however, the field still faces challenges, particularly in terms of adapting to variable environmental conditions (Blasco and Aleixos, 2023).

In our research, we have applied deep learning in the domain of feature detection and classification. Our CNN models have shown promising results in classifying corn seed varieties and determining the ripeness of mulberries. There are still areas for improvement. A notable limitation is the adaptability of these models to changing environmental conditions, which can greatly influence classification accuracy in real-world agricultural scenarios. Factors like lighting variations, background differences, and natural changes in appearance due to growth conditions can significantly affect model performance. Future research should aim at enhancing the robustness and flexibility of these CNN models. This enhancement could involve training the models with more diverse datasets that encompass a broader spectrum of environmental conditions and various stages of crop development. Incorporating techniques like transfer learning or fine-tuning with datasets from different geographical regions and climate conditions could also be beneficial. Moreover, expanding the application of these models to a wider range of products varieties could significantly bolster their usefulness in agronomic practices. By tackling these challenges, future work can ensure that deep learning methods become more universally applicable and reliable, adapting seamlessly to the dynamic conditions encountered in real-world agriculture (Shaikh et al., 2022). This progression is vital for the advancement of agronomy, as it moves towards more technologically driven practices, enhancing efficiency, reducing manual labor, and ultimately contributing to more sustainable and precise agricultural methodologies.

6.3.2 Expanding Biomedical Image Analysis (RQ3)

In addressing this research question, we utilized deep learning models for the segmentation task. Our success in segmenting zebrafish larvae images using the innovative integration of the Transformer and U-Net model has marked a significant breakthrough in the field of biomedical image analysis. Existing research in this area, particularly regarding VAST (Vertebrate Automated Screening Technology) images of zebrafish, has predominantly relied on more traditional image processing techniques or earlier versions of neural networks. These conventional approaches often face challenges in dealing with the high complexity and variability inherent in VAST images (Chang et al., 2012). They tend to struggle with accurately segmenting intricate anatomical structures and may lack the precision needed for detailed analysis.

Looking forward, future work should aim to expand the application of our segmentation methodology to other model organisms and diverse biomedical imaging scenarios. This expansion requires the model to adapt to various anatomical structures and unique characteristics of different image types. For instance, applying our model to high-resolution zebrafish images or extending its capabilities to human tissue analysis could significantly bolster its applicability in medical diagnostics and research.

Adapting the model to high-resolution images will necessitate optimization to manage larger data sizes and intricate features (Sampath et al., 2021). In the realm of human tissue analysis, the model would need extensive training on a wide array of tissue types and pathological conditions. This might also involve modifications to its architecture to effectively handle the greater complexity and variability found in human tissues. Additionally, continuous retraining of the model with new data and refining its learning algorithms will be crucial to improving its accuracy and versatility.

By addressing these challenges and building upon current research, we can significantly advance the utility of deep learning models in biomedical imaging. This advancement is not only crucial for improving medical diagnostics but also for providing deeper insights into biological processes, ultimately contributing to the advancement of medical science and healthcare.

6.3.3 Advancing Image Captioning Techniques (RQ4)

In addressing our fourth set of research questions, we employed deep learning models in conjunction with advanced Natural Language Processing (NLP) techniques, propelling our work in image captioning to new territory. Our advancements have laid a robust groundwork for future development, particularly in integrating more refined NLP methods. A current limitation of our model lies in its partial grasp of the broader context and emotional nuances in images. To overcome this, future research should focus on incorporating state-of-the-art NLP techniques, such as contextual understanding and sentiment analysis (Wang et al., 2023), (Gandhi et al., 2023). This could entail retraining our models to generate captions that capture a deeper level of detail and subtlety.

A crucial aspect of this future development involves training the model on a diverse array of image and text datasets. This approach would enhance the capacity of the model to interpret various contexts and emotions accurately. Moreover, exploring the integration of transformer-based language models, which have demonstrated remarkable success in grasping complex language structures and subtleties, could significantly improve the narrative depth of image descriptions (Cornia et al., 2022), (Salaberria et al., 2023), (Abdal Hafeth and Kollias, 2024). Implementing such advanced language processing techniques would not only refine the precision of image captioning but also broaden its applications. This technology could become increasingly relevant in areas such as automated content generation, assistive technologies for the visually impaired, and interactive AI systems, where accurate and context-aware image descriptions are crucial.

Furthermore, existing research in fields like automated journalism and social media content generation has begun to leverage similar technologies, indicating a growing trend in the application of AI-driven image captioning. By building upon these developments, future work in this area could lead to groundbreaking improvements in how AI understands and communicates visual information.

Overall, the research detailed in this thesis has laid a firm foundation for the application of deep learning across various image processing domains, including feature extraction, image classification, segmentation tasks, and integration with natural language processing. Looking ahead, the integration of these methods with emerging technologies such as augmented reality, robotics, and the Internet of Things (IoT) holds the potential to bring transformative changes. These advancements could profoundly impact sectors from smart technology in the postharvest agriculture industry, where AI can enhance precision farming techniques, to advanced healthcare systems, where improved image analysis can aid in diagnostics and treatment planning. Moreover, the incorporation of deep learning in preclinical research, particularly in drug discovery and development, presents exciting possibilities. The scope for interdisciplinary research is vast, opening a palette of opportunities to push the boundaries in Artificial Intelligence, Deep Learning, image analysis, and beyond.

As we conclude this thesis, we hope that the research presented herein contributes to the academic discourse as well as inspires practical applications and further explorations that continue to harness the power of artificial intelligence in solving real-world problems.