

The Disease of the Canine Eye - From Image to Diagnosis Using AI

Matija Buric

[0000-0003-3528-7550]

Marina Ivacic-Kos

[0000-0002-1940-5089]

Sanda Martincic-Ipsic

[0000-0002-1900-5333]

Faculty of Informatics and Digital Technologies and Centre for Artificial Intelligence, University of Rijeka, Ul. Radmile Matejcic 2,

51000 Rijeka, Croatia

matija.buric@uniri.hr

marinai@uniri.hr

smart@uniri.hr

ABSTRACT

This research examines the application of computer vision (CV) and large language models (LLM) in diagnosing eye diseases in dogs. The study utilizes a U-Net framework, incorporating convolutional neural networks (CNNs) such as ResNet, Inception, VGG, and EfficientNet, to enhance the segmentation of eye disease areas. Along the base U-Net model, four U-Net-based models were developed and evaluated on a dataset specifically generated for this purpose, classifying eye diseases into four categories. The performance of the enhanced U-Net architectures was found to be superior to that of the standard U-Net, with the U-Net modified with ResNet34 achieving the best segmentation accuracy, as measured by a Jaccard index of 66.6% on a custom test set. The segmented images were then diagnosed using various LLMs, including ChatGPT, Mistral, Gemini (Bard), Claude, and Llama-2, which were assessed using 15 different symptom sets. The study demonstrates that combining advanced image segmentation techniques with LLMs can improve diagnostic accuracy in veterinary medicine. The approach leverages the segmentation capabilities of U-Net for precise localization and the diagnostic ability of LLMs to interpret symptoms, facilitating enhanced diagnostic tools. This method could be applicable to other medical diagnostic areas requiring similar dual capabilities.

Keywords

Computer Vision (CV), Large Language Models (LLM), Image Segmentation, U-Net Architecture, Veterinary Ophthalmology, Disease Localization, Diagnostic Tools, Veterinary Diagnostic Imaging, Automated Medical Diagnosis.

1. INTRODUCTION

Major advances in computer vision and deep learning make these technologies more interesting for integration in veterinary medicine to facilitate early and accurate diagnosis to improve treatment outcomes [RoFB15].

The article discusses the development and performance of models trained for semantic segmentation of ocular symptoms in dogs with standard U-Net encoder-decoder network and enhanced with different CNN backbones. A specially generated dataset, featuring symptoms like corneal cloudiness and scleral redness, supports the training of our models [GDLM20]. For the interpretation of detected symptoms and the support of the diagnostic processes the LLM chatbots ChatGPT [SPKS24], Mistral [JSMB23] and Llama-2 [TMSA23] are evaluated. Gemini (used to be called Bard) [HMHS23], and Claude [WKBB23].

2. RELATED WORK

The CNNs have significantly advanced the medical diagnosis, with U-Net architecture emerging as particularly effective for medical imaging analysis due

to its robust feature extraction and segmentation capabilities [RoFB15]. Enhancements to this model include integration with advanced CNN backbones like ResNet [HZRS15] and Inception [SVIS15], improving symptom localization in various medical contexts [ChCh19, LXWJ19]. Competitive performance of U-Net-based models in medicine is demonstrated also in [AARJ22, PTRS21] and [JBZL22] that has shown that U-Net can outperform transformer-based architectures in medical image registration with minimal modifications. Research on canine ocular conditions remains limited, focusing primarily on diseases such as glaucoma [JoMK06, SHIS11]. The scarcity of datasets poses significant challenges, and while synthetic datasets have proven beneficial in other areas, they are less suitable for this field due to the variability in clinical presentations [DKKC21]. The efficacy of LLMs in medical settings

has been less explored, with existing methodologies for performance measurement often not meeting the complex requirements of healthcare interactions [ARKH23, DaZB18, WKBB23]. This research incorporates LLM technology with image segmentation to investigate their combined potential in enhancing diagnostic accuracy for canine eye diseases.

3. U-NET MODELS FOR CANINE OPHTHALMOLOGIC SYMPTOMS SEMANTIC SEGMENTATION

The diversity of eye diseases in dogs poses significant diagnostic challenges. For more effective classification, diseases were categorized into prevalent symptom groups to better focus the model on essential visual indicators. This categorization spans various conditions, enhancing the model's capability to generalize across different canine breeds and ocular pathologies.

Dataset Description and Augmentation

A custom dataset comprising 145 images sized 320x320 pixels was developed, sourced from specialized ophthalmology clinics, and reviewed by veterinary specialists. It represents diverse breeds and disease manifestations. Each image is paired with a uniquely colored mask highlighting four key symptoms, maintaining proportionate representation across different classes. Augmentation techniques like horizontal flips, rotations, and translations were employed to introduce variability and improve generalization; however, zooming was avoided to preserve mask integrity. The final dataset now consists of 200 images with distribution of symptoms appearances from 118 for excessive tearing to 154 for sclera redness because of multi-symptom occurrence on images.

U-Net Architectural Enhancements and Performance Evaluation

The U-Net framework was enhanced with various backbone networks to reinforce feature extraction. Configurations included ResNet34, known for its deep residual learning; Inception V3, which utilizes inception modules for efficient cross-channel correlation processing; VGG16, noted for its deep, straightforward convolutional structure [SiZi15], and EfficientNet B3, optimizing the balance among depth, width, and resolution [TaLe20]. Performance evaluations focused on the mean Jaccard index across different ocular conditions as detailed in Table 1. The U-Net with ResNet34 backbone exhibited the best performance in 3 out of 4 symptom segmentation accuracy tests. These modifications to U-Net architecture underscore its potential to enhance disease segmentation in veterinary ophthalmology.

Architecture	Ocular Opacity	Sclera Redness	Excessive Tearing	Colored Ocular Protrusion
U-Net	38.5	44	0.3	55.8
U-Net + ResNet34	73.9	80.6	38	73.9
U-Net + Inception V3	78.3	78.3	37.9	54.2
U-Net + VGG16	75.1	75.7	27.2	54.1
U-Net + EfficientNet	69.7	79.2	36.1	67.5

Table 1. Results of U-Net segmentation in Jaccard coefficient index showing performance using different backbones.

4. INTERPRETATION OF SYMPTOMS USING LARGE LANGUAGE MODELS

The integration of U-Net architecture in this experiment assures precise localization of ocular symptoms in dogs. This segmentation information, in the form of symptom presence, are fed into the diagnostic phase, where large language models (LLMs) like ChatGPT, Mistral, Gemini, Llama, and Claude analyze and interpret these symptoms, providing preliminary diagnoses. This dual approach combines U-Net's accurate localization with the analytical capabilities of LLMs to enhance the diagnostic accuracy and efficiency.

Evaluating Symptom Interpretation Using LLMs

To evaluate the LLMs performance in interpreting textual descriptions of disease symptoms, it was necessary to define an appropriate metric to measure the success of the model. Textual descriptions of symptoms are subjective, often containing ambiguous or incomplete information. Therefore, the evaluation metric focuses on effectiveness of LLMs to accurately interpret these descriptions and to extract relevant medical information and provide accurate, coherent responses that are consistent with medical knowledge.

To ensure the reliability of symptom interpretation, only segmentation results with an accuracy above 98% are considered. This strict threshold ensures that errors in segmentation do not compromise the diagnostic accuracy of the LLMs.

Comparative Evaluation of LLM Diagnoses based on U-Net Symptoms in Canine Eye Diseases

The study assessed the diagnostic accuracy of LLMs using 15 medical cases with varying symptom combinations, each derived from U-Net identified symptoms. The accuracy of the LLMs diagnoses was

evaluated using a straightforward scoring system which was based on expert knowledge for each individual case: a score of 1 was given for a correct match, 0 for indeterminate responses, and -1 for incorrect or hallucinated diagnoses. An example of such evaluation on an image taken from the [ThMP00] show an American bulldog with protrusion of pink colored mass and epiphora symptoms. These two symptoms were taken as a reference in setting the initial diagnosis which was presented to the LLM models. The prompt directed to the LLM was divided into context, desired output and query as shown in the Table 2. The analysis of LLM performance in diagnosing canine eye diseases highlights significant insights into their diagnostic capabilities. ChatGPT4 demonstrates the strongest overall performance, particularly in accurately diagnosing single-symptom conditions such as Nuclear Sclerosis and Cherry Eye. However, it, along with other models, encounters difficulties with complex multi-symptom scenarios. ChatGPT3.5 shows variability, effectively diagnosing conditions with complete symptom presence but struggling in their absence. Gemini and Claude are effective in straightforward cases but face limitations in complex scenarios. Mistral AI shows steadiness in simple cases, cautiously avoiding overconfident errors in more complicated diagnoses. Llama-2 struggles across a range of scenarios and fails to recognize excessive tearing as a symptom.

5. CONCLUSION

This study indicates a promising future in veterinary ophthalmology through the integration of computer vision and LLM technologies.

The implementation of U-Net particularly enhanced with backbones like ResNet34 and EfficientNet B3, has effectively segmented ocular symptoms in canines. The U-Net with ResNet34 notably outperformed the standard model, especially effective in recognizing corneal cloudiness and scleral redness. Challenges remain in segmenting excessive tearing due to difficulties in annotation and boundary determination. The development of a specialized dataset was crucial for optimizing the U-Net models to detect ocular symptoms accurately.

The utilization of LLMs such as ChatGPT, Mistral, Gemini, Llama-2, and Claude for symptom interpretation and preliminary diagnoses revealed that while LLMs manage single-symptom cases effectively, their performance decreases with the complexity of multiple symptoms. This observation underscores the need for further enhancement of AI diagnostic tools but also suggest potential in supporting diagnosis.

6. ACKNOWLEDGMENTS

This research was supported by an Image-based AI-assisted diagnostics of canine ocular disease (AICODD) project funded by Animal Eye Consultants of Iowa, Iowa, USA.

LLM	response to a Prompt*	Score
Llama-2	<i>Excessive Tearing + Coloured Mass Protrusion in the Eye Corner Diagnosis: Allergic Reaction</i>	-1
Mistral	<i>Excessive tearing or Watery eyes + Coloured Swelling in the Eye Edge: Although less commonly seen together, coexisting watery eyes and coloured swelling may warrant referral to a specialized veterinarian for proper evaluation and management of suspected immune-mediated disorders affecting multiple organs.</i>	0
GPT4	<i>Watery Eyes + Protrusion of a Red or Pink Mass - Diagnosis: Cherry Eye with secondary tear duct irritation.</i>	1
Claude	<i>Based on the symptoms you described, I do not have enough information to make a definitive diagnosis. The tearing and protruding mass indicate an abnormality affecting the eye and surrounding structures. But the specific diagnosis cannot be made based on these limited symptoms alone. Further veterinary examination of the eye would be recommended.</i>	0
Gemini	<i>The combination of excessive tearing and a red/pink mass protruding from the eye corner strongly suggests cherry eye, a prolapse of the third eyelid gland. This condition usually requires veterinary examination and potentially surgery for correction.</i>	1

* [Can you make a short pinpoint diagnosis based on the combination of these dog eye symptoms to assist a vet in preliminary decision-making for diagnosis:

1. Excessive Tearing/Watery Eyes
2. Protrusion of a Red or Pink Mass in the Eye Corner]

Table 2. Examples of prompt presented to three LLM models, along with their responses and associated score values. The Correct diagnosis is Cherry Eye.

7. REFERENCES

- [AARJ22] Azad, Reza, Aghdam, Ehsan Khodapanah, Rauland, Amelie, Jia, Yiwei, Avval, Atlas Haddadi, Bozorgpour, Afshin, Karimijafarbigloo, Sanaz, Cohen, Joseph Paul, et al.: Medical Image Segmentation Review: The success of U-Net, [object Object] (2022)
- [ARKH23] Ahmed, Imtiaz, Roy, Ayon, Kajol, Mashrafi, Hasan, Uzma, Datta, Partha Protim, Reza, Md. Rokonzaman: *ChatGPT vs. Bard: A Comparative Study* (preprint) : Preprints, 2023
- [ChCh19] Chakrabarty, Navoneel, Chatterjee, Subhrasankar: A Novel Approach to Glaucoma Screening using Computer Vision. In: *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. Tirunelveli, India : IEEE, 2019 — ISBN 978-1-72812-119-2, pp. 881–884
- [DaZB18] Dandekar, Ashish, Zen, Remmy A. M., Bressan, Stéphane: A Comparative Study of Synthetic Dataset Generation Techniques. In: Hartmann, S., Ma, H., Hameurlain, A., Pernul, G., Wagner, R. R. (eds.): *Database and Expert Systems Applications, Lecture Notes in Computer Science*. Cham : Springer International Publishing, 2018 — ISBN 978-3-319-98812-2, pp. 387–395
- [DKKC21] Deane, Jake, Kearney, Sinead, Kim, Kwang In, Cosker, Darren: DynaDog+T: A Parametric Animal Model for Synthetic Canine Image Generation, arXiv (2021). — arXiv:2107.07330 [cs]
- [GDLM20] Grozdanić, Siniša, Đukić, Slavica, Luzhetskij, Sergei, Milčić-Matić, Natalija, Lazić, Tatjana: *Atlas bolesti oka pasa i mačaka*. Beograd : Oculus Vet, 2020
- [HMHS23] Hirose, Takanobu, Mizuta, Kazuya, Harada, Yukinori, Shimizu, Taro: Comparative Evaluation of Diagnostic Accuracy Between Google Bard and Physicians. In: *The American Journal of Medicine* vol. 136 (2023), Nr. 11, pp. 1119-1123.e18
- [JBZL22] Jia, Xi, Bartlett, Joseph, Zhang, Tianyang, Lu, Wenqi, Qiu, Zhaowen, Duan, Jinming: U-Net vs Transformer: Is U-Net Outdated in Medical Image Registration? In: Lian, C., Cao, X., Rekik, I., Xu, X., Cui, Z. (eds.): *Machine Learning in Medical Imaging, Lecture Notes in Computer Science*. vol. 13583. Cham : Springer Nature Switzerland, 2022 — ISBN 978-3-031-21013-6, pp. 151–160
- [JoMK06] Johnsen, Devin A. J., Maggs, David J., Kass, Philip H.: Evaluation of risk factors for development of secondary glaucoma in dogs: 156 cases (1999–2004). In: *Journal of the American Veterinary Medical Association* vol. 229. Schaumburg, IL, USA, American Veterinary Medical Association (2006), Nr. 8, pp. 1270–1274
- [JSMB23] Jiang, Albert Q., Sablayrolles, Alexandre, Mensch, Arthur, Bamford, Chris, Chaplot, Devendra Singh, Casas, Diego de las, Bressand, Florian, Lengyel, Gianna, et al.: Mistral 7B, arXiv (2023)
- [LXWJ19] Li, Liu, Xu, Mai, Wang, Xiaofei, Jiang, Lai, Liu, Hanruo: Attention Based Glaucoma Detection: A Large-Scale Database and CNN Model. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA : IEEE, 2019 — ISBN 978-1-72813-293-8, pp. 10563–10572
- [PTRS21] Petit, Olivier, Thome, Nicolas, Rambour, Clément, Soler, Luc: U-Net Transformer: Self and Cross Attention for Medical Image Segmentation, [object Object] (2021)
- [RoFB15] Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv (2015). — arXiv:1505.04597 [cs]
- [SHIS11] Strom, Ann Refstrup, Hässig, Michael, Iburg, Tine M., Spiess, Bernhard M.: Epidemiology of canine glaucoma presented to University of Zurich from 1995 to 2009. Part 1: Congenital and primary glaucoma (4 and 123 cases). In: *Veterinary Ophthalmology* vol. 14 (2011), Nr. 2, pp. 121–126
- [SiZi15] Simonyan, Karen, Zisserman, Andrew: Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv (2015). — arXiv:1409.1556 [cs]
- [SPKS24] Savage, Cody H., Park, Hyungsun, Kwak, Kijung, Smith, Andrew D., Rothenberg, Steven A., Parekh, Vishwa S., Doo, Florence X., Yi, Paul H.: General-Purpose Large Language Models Versus a Domain-Specific Natural Language Processing Tool for Label Extraction From Chest Radiograph Reports. In: *American Journal of Roentgenology* (2024), p. AJR.23.30573
- [TaLe20] Tan, Mingxing, Le, Quoc V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, arXiv (2020). — arXiv:1905.11946 [cs, stat]
- [ThMP00] Thamizharasan, A, Murugan, M S, Parthiban, S: Surgical Management of Cherry Eye in a Dog
- [TMSA23] Touvron, Hugo, Martin, Louis, Stone, Kevin, Albert, Peter, Almahairi, Amjad, Babaei, Yasmine, Bashlykov, Nikolay, Batra, Soumya, et al.: Llama 2: Open Foundation and Fine-Tuned Chat Models, arXiv (2023). — arXiv:2307.09288 [cs]
- [WKBB23] Wu, Sean, Koo, Michael, Blum, Lesley, Black, Andy, Kao, Liyo, Scalzo, Fabien, Kurtz, Ira: A Comparative Study of Open-Source Large Language Models, GPT-4 and Claude 2: Multiple-Choice Test Taking in Nephrology, arXiv (2023)