Research Proposal: OpenAI APIs for Dermatology Image Captioning

Madison Hubley, Mohd Faizan, David Awodiji, Liyang Wan

Dalhousie University

CSCI 6410/4148 / EPAH 6410: Applied Research in Health Data Science

Dr. Finlay Maguire

June 21, 2024

Contents

Abstract

The increasing incidence of skin cancer and other dermatological conditions needs more efficient diagnostic tools to improve patient outcomes. The project defines an approach to generate detailed captions for dermatological images, assisting in the diagnosis and treatment of skin lesions. To achieve this objective, we aim to use a technique that combines retrieval and foundational models to generate detailed captions for dermatological images. This dual-model approach utilizes integrating CLIP (Contrastive Language-Image Pre-training) for visual feature extraction and BERT (Bidirectional Encoder Representations from Transformers) for generating text captions. The training will employ a diverse dataset including images annotated with detailed clinical descriptions and demographic from the SCIN dataset by Google Research. To further improve the accuracy, relevance and reduce hallucinations in our model's outputs we implement the Retrieval-Augmented Generation (RAG) framework which will use a knowledge base constructed from prevalent dermatological datasets, including images and their annotations (DermNet, MedNode, and PAD-UFES-20). It will serve as a digital assistant to dermatologists, reducing their workload and enabling them to make more informed decisions, improving patient outcomes through earlier and more precise treatment interventions. Evaluation metrics include precision, recall, F1-score, semantic coherence (assessed through BLEU and ROUGE metrics), and clinical relevance, validated by dermatological experts. The study also involves user acceptance and usability testing, with iterative feedback loops from dermatologists to refine the model. Ethical considerations such as bias mitigation, data privacy, transparency, and compliance with regulations are addressed to ensure the responsible deployment of the technology.

Lay Summary

Due to the high mortality rate of many skin conditions, particularly melanoma, in combination with the incredibly high workload placed on medical specialists, now is the ideal time to research and develop new medical tools which can assist in diagnostics and speed up the treatment process. Dermatological screenings and related tests are time consuming procedures which impose a burden on not only professionals, but patients as well. With the recent improvements in the field of artificial intelligence (AI), there is greater potential now than ever in using machine learning models to assist in the analysis of skin lesions. By creating a tool that is able to not only classify images of skin lesions, but provide details about specific patterns and details, medical specialists will essentially have a trained personal assistant by their side. The proposed research aims to train a machine learning model to generate captions for images of skin lesions, describing features of the image that will assist in diagnostics. The use of OpenAI’s APIs implements cutting-edge technologies to assist in text-generation for the captioning process, creating detailed and coherent descriptions of what the model sees. The implementation of such a tool in dermatology has the power to detect nuances that may go undetected by even a trained

human eye, thus speeding up the process of diagnostics and allowing for earlier treatments, consequently improving patient outcome.

Introduction

Routine dermatology screenings often include checks for certain skin lesions or areas of concern, and if necessary, further measures to ensure that these spots are not dangerous to the individual. The first step of the screening is typically a visual observation by a medical professional, and if the spot is believed to be of concern, tests will be performed to further investigate. These tests, however, are both expensive and time-consuming, which can easily result in long wait times and a backlog of patients. With the high workload of medical professionals, and in an attempt to not overwhelm patients who likely do not have serious conditions, it makes sense to visually identify lesions of concern while disregarding those that appear to be harmless.

Identifying “harmless” lesions, however, is incredibly difficult. Medical professionals, despite their rigorous training, are still able to misdiagnose or make mistakes, especially when there is such a wide variety of appearances that dangerous skin lesions can take on. In addition to this, human skin is far from uniform. If a student is only trained on textbooks showcasing lesions on one skin color or type, they will likely have difficulty identifying the same lesion if it presents differently on another skin type. Thus, the significance of this primary visual observation should not be understated. Medical professionals receive extensive training to be able to identify areas of concern, but the possibility of error still exists, especially for those with less practical experience.

As observed in the 2008 paper by Heal et al., “accuracy of diagnosis [of skin lesions] was dependent on body site”, indicating that errors are more common on particular parts of the body, likely where lesions are less commonly located. Azeem et al. note that “while melanoma accounts for 4% of skin cancer cases, it causes 75% of skin-cancer-related deaths," a tremendously high statistic (2023). If a false negative occurs, patients will suffer as a result. This is especially detrimental if the patient does not book a follow up, as they may not have another chance to catch an illness. Thus, it is of the utmost importance to patient care to prevent false negatives, as there is no guarantee that there will be another opportunity to fix these mistakes.

The use of computational tools in medical care is not limited to dermatology - robotic surgery is a relatively new, yet widespread application of this, which has “improved patient outcomes... and enhanced surgical precision” (Reddy et al., 2023). With the recent surge of improvements in artificial intelligence, the integration of AI in medicine has seen an increase in research, including in the field of dermatology specifically.

Amongst this, a common area of interest is the potential usage of neural networks to classify images of skin lesions. If advances are to be made in dermatology, the inclusion of computational tools such as neural networks are an objectively promising way to go. Having a specially trained AI tool to analyze skin lesions or patches would be incredibly beneficial to dermatologists. Not only could this help draw attention to subtle patterns a human may not initially see, but it can increase efficiency and serve as an assistant. It should be noted that no tool should seek to replace medical professionals, but rather assist them in order to benefit their patients.

Literature Review

Along the recent surge in AI advancements, we are also seeing a surge in research on the plausibility of using machine learning and artificial intelligence in the medical field - specifically in dermatology. The literature expresses not only a need for new technologies to aid in the detection of abnormal skin lesions, but a promising outlook for the machine learning technologies which have been developed to do so.

Numerous machine learning models have been developed and trained to classify skin lesions, particularly with a focus on skin cancers. Much of the research on this matter focuses on convolutional neural networks due to their promising accuracy in said application. In the work by Shetty et al., researchers “obtained an accuracy of 95.18% with the CNN model” they had trained to classify images of lesions into one of seven categories (2022). Of similar success, Azeem et al. compared a number of previously implemented “CNN models for skin lesions detection on different dermoscopy datasets”, all of which obtained accuracies of at least 90% (2024). Of these CNN models, the one which had the lowest accuracy of 90% was a CNN utilizing the HAM10000 dataset, where researchers noted that tuning hyperparameters would be needed to increase the model’s accuracy (Azeem et al., 2024). The model with the highest accuracy of 97.49% was a CNN utilizing the ISIC dataset, and although the results were very impressive, it was noted that the quantity of images in the dataset was lacking (Azeem et al., 2024).

One persistent issue in the literature involves datasets; even in some of the most widely regarded datasets, data imbalance proves to be a notable concern, as well as a “significant challenge” to research wishing to utilize them, according to Shetty et al. regarding their usage of the HAM10000 dataset (2022). Datasets which do not accurately represent the variety of skin lesions experienced in a clinical setting in their respective proportions will negatively influence the success of any model which utilizes them for training, thus limiting the model’s potential applications. Training machine learning models to recognize a larger variety of skin lesions beyond just cancers requires a much greater amount of data, time, and computational power. On top of this, it requires care that the datasets have a wide variety of images, skin tones, and skin types. Thus, it is no surprise that there is a lack of research on the use of AI to perform dermatology image captioning. The success in AI classification of lesions is promising, yet the tremendous amount of detail provided by image captioning is an even more valuable asset to healthcare practitioners, and of course, patients. Captions noting subtleties of lesions which humans may not pick up on has the potential to lead to better diagnoses, improving the rate of treatment and overall patient outlook - something which there is an undeniable need for, and which the literature greatly supports. With the many recent advancements in artificial intelligence, now is a better time than ever for such an idea to be researched.

Much of the literature on dermatological care emphasizes the necessity of early detection, and the potential of machine learning algorithms to help solve this problem. Bhatt et al. reference findings that “the timely identification of early-stage skin cancer reduced the mortality rate by 90%”, something essential to consider since “melanoma is the fifth most common invasive cancer in the USA, and its incidence is increasing around the world” (Chan et al., 2020). Additionally, the most common way to diagnose skin cancer, biopsies, are time consuming and expensive (Chen et al., 2023), it is not uncommon for specialists to find themselves with a patient backlog. Likely influenced by the long wait times and backlogs, “data from the National Health Interview Survey indicate that screening rates are remarkably low (16% in men and 13% in women)” (Chan et al., 2020). Many papers express the sentiment that machine learning algorithms have the potential to help in this regard, “lowering the workload of specialists while simultaneously enhancing skin lesion diagnostics” (Bhatt et al., 2023). The assistance of a tool which is not only capable of recognizing concerning features of skin lesions, but also providing captions with detail, would be of tremendous benefit to healthcare professionals. Screenings would be sped up and have the assistance of AI, leading to official diagnoses being made sooner and with more accuracy. The reduction of wait times could result in an increased percentage of individuals willing to undergo dermatological screenings, thus increasing the odds of diseases being diagnosed early and improving patient outcomes.

Methodology

While ML models are frequently developed in general image captioning and broader medical imaging contexts, their application in dermatology remains unexplored. The work defines an approach to generate detailed captions for dermatological images, assisting in the diagnosis and treatment of skin lesions.

Primary Research Question: How can a machine learning approach combining retrieval and foundational models to generate detailed captions for dermatological images by integrating visual and textual analysis improve the accuracy and detail of dermatology image captions sufficiently to aid in early disease detection and diagnosis?

Secondary Research Questions: How does the Retrieval-Augmented Generation (RAG) framework influence the performance and reliability of AI-generated dermatological captions in reflecting true clinical conditions?

Data collection

The study will use a dataset of dermatology images with corresponding detailed clinical descriptions. The stills will be drawn from the 'SCIN' dataset by Google Research: a collection of 5,000+ volunteer contributions (10,000+ images) of common dermatology conditions. The individual contributions are described, with self-reported demographic, history, symptom information, self-reported Fitzpatrick skin type (sFST), dermatologist labels of the skin condition and estimated Fitzpatrick skin type (eFST) and layperson estimated Monk Skin tone (eMST) labels being provided for each contribution (Ward et al., 2024). This particular dataset is unique in its depth - the diversity of people and skin conditions represented is important to make sure the AI model works well across a wide range of skin types and diseases. This included 35.0% who identified as White, 5.3% as Black or African American, 4.5% as Hispanic Latino or Spanish Origin, 1.7% as Asian and 1.0% as American Indian or Alaska Native; apart from those mentioned less than 1% identified across other categories, while for (0.7%) they preferred not to answer.

Data Preprocessing

The data will be normalized into a similar scale and intensity, resizing images into specific dimensions to convert them into a common size, and the color adjustment to enhance the model accuracy under various conditions. When applicable, natural language editing will be performed to remove stains, standardize medical vocabulary, and encode for compatibility with models while maintaining semantic cleanliness. We plan to write pipelines in such a way as to have the ability to automatically preprocess data in a scalable and reproducible manner, so that any step can be picked up by another team in future research.

Variable Selection

Image Feature: We use pixel intensity, color histograms, texture patterns, and edge features for capturing the visual signatures of skin conditions. Pixel intensity and color histograms capture the visual characteristics of color distribution (which is especially important for distinguishing different skin conditions based on visible symptoms like color variation), serving as initial classification features. Texture Patterns and Edge Features help differentiate between lesions that are of similar colors but have varying surface structures; necessary in diseases such as eczema or psoriasis. Integration for defining lesion borders, crucial for malignant skin lesions like melanoma diagnosis, depends on edge detection.

Annotation Features: Text annotating lesion appearance, location, and possibly progression over time will be used for training the NLP components of the model. A new lesion appearance description layer will detail the visual aspects of lesions suspected by dermatologists for an initial diagnosis. In more complex terms, this means the model will be trained to generate and recognize descriptive text based on these annotations, resulting in the ability to provide relevant captions that are of clinical importance that could be flagging some skin conditions based on body location information. We will use annotations in describing longitudinal data that includes other lesions pertaining to the lesion and its progression over time to help support model training for patterns of progression that are suggestive of malignant or benign conditions.

Demographic Variables: Including age, skin type, and ethnic background will fine-tune the AI model to detect multiple variations of skin conditions based on their demographic-specific prevalence and presentation. The addition of age will assist in predicting age-related skin disorders, whereas skin type and ethnicity are crucial for correct diagnosis in diverse populations, making the model effective and avoiding bias in clinical practice.

Model Selection Rationale and Training

We will be using CLIP and BERT for this research because of their capabilities in processing image and textual data, respectively. Employing a dual-model approach integrating CLIP for visual feature extraction and BERT for generating text captions, augmented by a Retrieval-Augmented Generation (RAG) framework to utilize a knowledge base during the generation process.

CLIP (Contrastive Language-Image Pre-training): Its ability to understand and encode images in the context of natural language. Experiments demonstrate the superior suitability of CLIP-based encoders compared to non-CLIP-based, for multi-modal tasks such as image captioning (Barraco et al., n.d.). This model excels in identifying nuanced visual features across varied images, making it ideal for the initial interpretation of dermatological images. The model will be trained on dermatological images and corresponding textual descriptions. Each image is paired with a text snippet that describes its content. It learns from these pairs by maximizing the cosine similarity between the embeddings of correctly matched image-text pairs, a process that strengthens its capability to extract and interpret relevant visual features accurately.

BERT (Bidirectional Encoder Representations from Transformers): Its advanced capabilities in natural language processing, specifically its effectiveness in generating contextually rich textual outputs. BERT's ability to understand and generate detailed medical descriptions based on image features makes it suitable for creating accurate and informative image captions. It employs a supervised learning approach using the captions generated by CLIP as initial inputs, refined by the detailed annotations available in the dataset. The data handling is done by providing BERT with textual descriptions that have been cleaned, encoded, and structured to align with the outputs from the CLIP model.

Knowledge Base Construction and RAG Framework Integration

Knowledge Base Development: A main aspect of this project is the formation of a comprehensive knowledge base, which is planned to be constructed from three critically acclaimed resources:

- **DermNet Dataset**: With over 23,000 images with detailed clinical annotations, this dataset offers a vast spectrum of dermatological conditions.
- **MedNode Dataset**: Comprising 10,000 images, this dataset enriches the knowledge base with specific disease identifications and affected body areas, crucial for localized diagnosis.
- **PAD-UFES-20 Dataset**: Specializing in pigmented skin lesions, this collection of 2,298 images provides essential data on pigmentation disorders, augmenting the depth and scope of the knowledge base.

Redundancy Detection: To determine redundancy in our datasets, we will employ feature similarity metrics and clustering algorithms. By identifying highly similar images and annotations, we can minimize redundancy and ensure a diverse training set that enhances the model's generalization capability.

RAG Model Implementation

To fully leverage the developed knowledge base, we propose the implementation of Retrieval-Augmented Generation framework that utilizes a dual-encoder framework.

API: A foundational model's API will be used to facilitate the generation of accurate captions. This API will integrate seamlessly with platforms such as OpenAI API, Hugging Face's Transformers, Google Cloud AI and Vertex AI utilizing their robust computational infrastructure to enhance processing capabilities.

Feature-Based Retrieval: By employing BERT and CLIP, the model will extract pertinent features from dermatological images and use these features to query the knowledge base, retrieving information that matches the visual indicators present in the images.

Caption Synthesis: The retrieved information will then be synthesized by the RAG model to generate detailed captions that not only describe the images accurately but also include diagnostic insights and potential treatment options.

Evaluation Metrics

Including precision, recall, and F1-score, these metrics will evaluate how precisely and completely the model identifies and describes dermatological conditions from images.

- **Semantic Coherence and Clinical Relevance:** To assess the linguistic quality and medical accuracy of the generated captions, semantic coherence will be quantitatively evaluated through metrics such as BLEU and ROUGE. Clinical relevance will be assessed through structured reviews by dermatological experts.
- **User Acceptance and Usability Testing:** Practical deployment scenarios will be simulated to collect qualitative feedback from end-users, primarily dermatologists, focusing on the model's integration into clinical workflows, ease of use, and its utility in enhancing diagnostic accuracy.

Cross-Validation

- **A/B Testing:** This will involve comparing the performance of the RAG model against established baseline models and manual annotations by dermatology experts. This comparative analysis helps benchmark the model's performance and gather actionable insights.
- **Iterative Feedback Loops:** We plan to engage dermatologists who will use the model in real-world settings. Their feedback will be instrumental in refining the model, focusing on enhancing the user experience and diagnostic accuracy.

Establishing a Baseline and Incorporating Feedback

We will establish baseline performance by initially training the model on a subset of the dataset and evaluating its performance using metrics like accuracy, precision, recall, and F1-score. Feedback from dermatologists will be incorporated iteratively. After each evaluation cycle, the model will be fine-tuned based on the expert feedback, improving its diagnostic accuracy and usability.

Budget and Timeline

Collecting skin images In order to teach AI to recognize skin cancer, we need lots of images of skin spots. The diverse collection should include different types of skin lesions, all colors and images from multiple angles. To collect this data set needs a lot of energy and resources. We need to cooperate with hospital and dermatology clinics, to provide compensation for the participation of patients, and ensure that the image collection meets consistent and high quality standards.

Teaching AI: We need a data scientist to fine-tune the progress of the algorithm and constantly check the AI study to ensure that it is effective. This step is very important to ensure that our AI tools are reliable.

Time: 6 months

Cost: $ 50,000 (Based on the survey results of the annual income of a data scientist in Canada, we made a proportional budget)

Hardware Purchase and Software Installations: Running AI models, especially those involving the image recognition model, needs powerful computers and special software. We need to invest in high performance servers and GPUs to handle large data sets and complex calculations. In addition, we need to get used to the promotion of machine learning and image processing software tool licenses. The infrastructure is to support the whole AI, the backbone of the project.

Cost: $ 5,000 (A high-end PC customized according to market price)

Real Life Environment Testing: It is important that once models have been properly trained it is tested in a real life scenario. We need to collaborate with hospitals, where we will integrate the AI systems into their work processes, and observe its performance on actual patients and uses with professionals. At this stage, our team and medical personnel will be closely monitoring the uses and results, to properly assess the accuracy and usability of the model in a clinical environment, feedback will be collected for any necessary improvements.

Time: 8 months

Cost: $100,000 (Based on a subsidy of $1,000 per patient)

Compliance with the regulations: In the medical field, abiding by the laws and regulations is very important. We must make sure that our project is in line with all the legal and moral standards, in order to protect the security and privacy of patients. This includes the patients' consent, to ensure data security and comply with HIPAA and other health care legislation. We may need to procure legal advisers to guide us through these laws and regulations, and ensure that every aspect of our project meets the requirements of the standard.

Time: 1 month

Cost: $ 19,000 (Based on the average hourly wage of Canadian professional lawyers of $76.86, we estimate that legal consultation on privacy protection will take 8 hours a day for 30 days)

Other important matters: Due to the nature of every venture there are always unforeseen circumstances or expenses that can occur, here we will allocate some part of the budget to mitigate these unforeseen instances and ensure we are not faced with any challenges without a proper plan or means to do so.

Cost: $ 20,000 (As a reserve of emergency funds, it will not be used easily)

Bottom line: In order to successfully develop and implement to help doctors detect cancer artificial intelligence systems, we need a total of $194,000 and 15 months. Our budget covers the training and testing of artificial intelligence, the necessary equipment and software needed, clinical trials, ensuring regulatory compliance, the implementation of our proposal as well as other basic expenses. Using this budget, we aim to create a reliable and effective tool to assist doctors in providing better care for patients.

Ethics

Are there any ethical implications of using such a technique?

The implementation of AI-generated captioning systems in dermatology raises several ethical implications that must be carefully considered and addressed. One of the primary ethical concerns is the potential for biases within the model. If the data used for training the AI system contains biases towards certain subgroups, these biases can be shown and to some degree amplified in the model's responses. This could lead to unfair treatment and disparities in its accuracy across different demographic groups. Ensuring diverse and representative datasets is essential to mitigate this risk.

Another significant ethical concern is privacy. The use of large datasets often contains sensitive personal information that poses considerable privacy risks. The datasets involved in this methodology may potentially include identifiable patient information, which if not properly safeguarded, could lead to breaches of privacy and confidentiality. Strong data protection measures, such as anonymization and encryption are needed to ensure that personal information remains secure and is not misused. Additionally, obtaining explicit user consent for data collection and usage, alongside clear explanations of how their data will be used in both research and monetization as well as how the data is protected, is crucial for maintaining trust and ethical integrity.

Transparency in the operation of AI systems is also a huge ethical issue. The technical and operational aspects of these systems are often not fully communicated to users, which leads to a lack of understanding and trust. Moreover, the absence of proper regulations and oversight can worsen these issues. It is important to provide clear and accessible information about how the technology works and the methods of user protection in place. Well detailed documentation that explains the model's operations, data sources, and processing methods will help build trust and understanding among users and collaborators.

How can this be done in a morally acceptable manner?

Ensuring the moral acceptability of this technique involves several key strategies. Models must first be designed with robust content moderation systems to filter out potentially harmful or inappropriate information. Regular updates to the knowledge base are essential to prevent the spread of outdated or incorrect information, thereby safeguarding user safety and well-being. Establishing accountability mechanisms such as audits and peer reviews is crucial for monitoring system performance and compliance to ethical standards. Implementing clear ethical guidelines and frameworks will define the responsible use of our methods, ensuring that all stakeholders understand and follow these standards.

However, there are potential push backs that must be addressed. One major concern is the discomfort some individuals may feel about having their data included in public or shared knowledge bases. Despite anonymization efforts, there remains a fear of potential data breaches or misuse. Addressing these concerns through rigorous data protection measures and clear communication is vital. Additionally, the methodology involved, particularly Retrieval-Augmented Generation (RAG), is not widely known or understood, this lack of familiarity can lead to reluctance and skepticism among users, shareholders, and practitioners. Education and additional outreach efforts are necessary to explain the benefits and uses of the technique, to reduce resistance and promote acceptance.

Finally, ensuring compliance with regulations such as the General Data Protection Regulation (GDPR) presents significant challenges. The use of data must align with various legal requirements, and any breaches, whether intentional or not, carry great legal risks and repercussions. Addressing these issues requires careful planning, legal oversight, and adherence to best practices in data management. By prioritizing bias mitigation, privacy protection, transparency, accountability, and user education, we can ensure that the deployment of this technology is both ethically sound and widely accepted by all.

Discussion

What are the limitations of this method?

One of the primary limitations of our proposed method is its complexity. The development and maintenance of a sophisticated AI-generated captioning system requires both a certain level of technical expertise and infrastructure. This complexity can pose challenges for both implementation and troubleshooting, which calls for ongoing support and refinement.

High resource demand is another significant limitation. The computational resources required for training and deploying AI models are substantial. This includes not only hardware such as powerful GPUs and large storage capacities but also the human resources needed to manage and optimize these systems.

The quality of data is a critical factor in the success of AI models. Our system's performance is heavily dependent on the quality and diversity of the data it is trained on. Any biases or inconsistencies in the training data can lead to inaccuracies and unintended consequences in the generated captions. This dependency underscores the importance of rigorous data analysis and preprocessing.

Biases are an inherent risk when dealing with AI systems. Despite our best efforts to use high-quality, representative data, there is always a chance that biases could be introduced during the training process. These biases could stem from various sources, including the training data selected, the labeling process, and the algorithms themselves. Addressing these biases and trying our best to limit any biases is essential to ensure fair and accurate results from the system.

What can be done to further improve or extend this methodology?

To further improve and extend this methodology, several strategies can be employed. Increasing knowledge about the methodology to the broader research and medical communities can foster better understanding and perhaps collaboration. Sharing insights and experiences can lead to the refinement of techniques and the identification of best practices and further drives helpful collaboration.

Technological advancements play a crucial role in enhancing AI systems. As technology evolves, improvements in computer systems, algorithms used, and tools at our disposal can greatly boost the efficiency and effectiveness of our approach. Fully taking advantage of these advancements that occur can greatly help overcome any current limitations.

Improving data quality and availability is another critical area for future work. By curating more diverse and high-quality datasets, we can enhance the robustness and accuracy of our AI models. Efforts to standardize data collection and labeling processes can also lead to better outcomes.

Ethical regulations are vital in guiding the responsible development and deployment of AI systems. Establishing and adhering to strict and vast ethical guidelines which cover a variety of different situations and bases help mitigate risks associated with bias, privacy, and misuse of AI.

Engaging with regulatory bodies and stakeholders to create comprehensive ethical frameworks is essential for the sustainable advancement and use of this technology.

What are the positive and negative implications of this proposal?

One of the significant positive implications is the potential for AI-generated image captioning to assist dermatologists in making more accurate and quicker diagnosis. By providing additional insights that may not be immediately apparent to human observers, the system can enhance diagnostic accuracy and efficiency. This, in turn, leads to more effective use of healthcare resources, quicker and more accurate diagnoses as well as better patient outcomes.

Furthermore, the integration of AI in diagnostic services can elevate the overall quality of healthcare. Enhanced diagnostic capabilities can contribute to improved public health by reducing the likelihood of misdiagnosis and enabling timely interventions. The widespread adoption of such technologies can drive innovation and set new standards in medical practice.

However, there are also potential negative impacts to consider. One such concern is the risk of over-reliance on AI. If clinicians become too dependent on the generated captions, there is a danger that their diagnostic skills and critical thinking abilities may diminish over time. This over-reliance could lead to complacency and a reduced ability to identify and address errors in AI outputs.

Moreover, the inherent uncertainties in AI systems mean that they can sometimes produce incorrect or misleading results. Blind trust in these systems without proper oversight and validation can result in adverse outcomes. It is crucial to highlight the complementary role of AI, ensuring that it is used as an aid rather than a replacement for human expertise.

Knowledge Translation

How can we make it replicable and easily accessible to others?

To promote transparency and reproducibility, all source code, scripts, and documentation used in our research will be made available on GitHub. The repository will be meticulously documented, including comments within the code and comprehensive readme files. These readme files will provide step-by-step instructions for setting up a testing environment, running the codebase, and reproducing the results. This approach ensures that other researchers and developers can easily replicate our work, verify our results, and build upon our methods.

The datasets utilized in our research are publicly accessible. Detailed instructions on how to access these datasets, along with their sources, will be provided. Additionally, we will document the preprocessing steps and any transformations applied to the data during the research process.

This transparency is crucial for reproducibility and allows other researchers to understand and replicate the exact conditions under which our results were obtained.

Our research paper will also be published in open-access journals and repositories, ensuring that it is widely accessible to those interested in the subject. The publication will include all necessary details and supplementary materials required for reproducibility. By making our findings freely available, we aim to foster greater collaboration and innovation within the research community.

How do we plan to share our findings across multiple platforms/mediums?
We will actively engage with the research community through various collaborative platforms such as forums, mailing lists, and social media. This engagement is intended to encourage collaboration, feedback, and adoption of our methods. By facilitating discussions and addressing queries, we hope to enable other researchers to recreate our results or explore different outcomes. To also further disseminate our findings, we plan to host workshops and webinars. These events will serve as platforms to present our research, demonstrate the tools and techniques used, and answer questions from the community. Workshops provide an interactive environment where participants can gain hands-on experience and foster a deeper understanding of our methods. Alongside our workshops an interactive website will be developed to showcase our findings in an engaging and accessible manner. This website will feature visualizations, data summaries, and interactive elements that allow users to explore the data and results in various ways. Such a platform not only enhances the accessibility of our research but also provides a novel way for users to interact with and understand the data.

What do we plan to do with our research?

One of the significant applications of our research is the creation of a library of dermatology cases annotated with AI-generated captions. This library can be utilized for educational purposes, serving as a reference for clinicians, students, and individuals in the dermatology field. By providing detailed annotations, we aim to enhance the learning experience and knowledge base in dermatology. The AI-generated captions developed through our research can also be used as a supplementary tool for clinicians, providing a second opinion to support their diagnoses. By offering additional analysis, the system can help confirm or refute initial diagnoses, thereby improving the accuracy and confidence of clinical decisions.

We plan to collaborate with healthcare technology companies to integrate our captioning system into dermatology-related products such as diagnostic tools, mobile apps, and electronic health record systems and additionally the system can be used to refute claims in dermatology insurance areas, providing a robust tool for various commercial applications. Before our collaboration we will firstly reach out to clinicians and healthcare admins proposing our methods detailing the advantages of using our work as well as it can be implemented. It is important to us that the clinicians know and view our proposal as only positives but we also want them to be comfortable in our proposal so it is important to note that this may not be possible if clinicians and healthcare admins are not comfortable using our proposed methods.

The AI system can be employed to provide consistent and standardized annotations for dermatology images, thereby improving the quality and reliability of image databases used for research and clinical purposes. This standardization is crucial for maintaining high standards in dermatological research and practice.

Our research will be accessible to the medical and academic community, including dermatologists, medical practitioners, researchers, and students. By sharing our findings with this community, we aim to advance knowledge in the field, foster collaboration, and inspire new research and innovations. Healthcare technology firms and AI-focused startups will have access to our research, enabling them to integrate our findings into their products and services. Companies developing medical software and diagnostic tools can build on our research to create innovative solutions and improve existing technologies. We also believe it crucial to make our research available to patients and the general public, by increasing awareness of dermatological conditions and providing information on early signs, we empower patients to take proactive steps in managing their health. Making this information publicly available can lead to better health outcomes and a more informed community.

References

Alaparthi, S., & Mishra, M. (2020). *Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. In arXiv [cs.CL].* http://arxiv.org/abs/2007.01127

Azeem, M., Kiani, K., Mansouri, T., & Topping, N. (2023). *SkinLesNet: Classification of Skin Lesions and Detection of Melanoma Cancer Using a Novel Multi-Layer Deep Convolutional Neural Network.* Cancers, 16(1), 108. https://doi.org/10.3390/cancers16010108

Barraco, M., Cornia, M., Cascianelli, S., Baraldi, L., & Cucchiara, R. (n.d.). *The unreasonable effectiveness of CLIP features for image captioning: An experimental analysis. Thecvf.com. Retrieved June 21, 2024, from* https://ieeexplore.ieee.org/document/9857436

Bhatt, H., Shah, K., Shah, M., Shah, R., & Shah, V. (2023). *State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review*. https://doi.org/10.1016/j.imed.2022.08.004

Chan, S., Reddy, V., Myers, B., Thibodeaux, Q., Brownstone, N., & Liao, W. (2020). *Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations. Dermatology and therapy*, 10(3), 365–386. https://doi.org/10.1007/s13555-020-00372-0

Chen, X., Chu, Y., Gao, X., He, X., Liao, X., Sun, L., Xu, J., Zhang, B., Zhou, L., & Zhou, J. (2023). *SkinGPT-4: An Interactive Dermatology Diagnostic System with Visual Large Language Model*. https://doi.org/10.48550/arXiv.2304.10691

DermNet. (2015, September 14). Dermnetnz.org; DermNet. https://dermnetnz.org/

Gomolin, A., Netchiporouk, E., Gniadecki, R., & Litvinov, I. V. (2020, March 5). *Artificial Intelligence Applications in dermatology: Where do we stand?*. Frontiers. https://www.frontiersin.org/articles/10.3389/fmed.2020.00100/full

Heal, C. F., Raasch, B. A., Buettner, P. G., & Weedon, D. (2008). *Accuracy of clinical diagnosis of skin lesions*. The British journal of dermatology, 159(3), 661–668. https://doi.org/10.1111/j.1365-2133.2008.08715.x

Huang, Y., & Huang, J. (2024). A survey on retrieval-augmented text generation for large language models. In arXiv [cs.IR]. http://arxiv.org/abs/2404.10981

Hyeon Ki Jeong 1 2, 1, 2, & Artificial intelligence (AI) has recently made great advances in image classification and malignancy prediction in the field of dermatology. (2022, August 23). *Deep learning in dermatology: A systematic review of current approaches, outcomes, and*

limitations. JID Innovations.
https://www.sciencedirect.com/science/article/pii/S2667026722000583

Liopyris, K., Gregoriou, S., Dias, J., & Stratigos, A. J. (2022). Artificial intelligence in dermatology: Challenges and perspectives. Dermatology and Therapy, 12(12), 2637–2651. https://doi.org/10.1007/s13555-022-00833-8

Li, Y., Liang, F., Zhao, L., Cui, Y., Ouyang, W., Shao, J., Yu, F., & Yan, J. (2021). Supervision exists everywhere: A data efficient Contrastive Language-Image Pre-training paradigm. In arXiv [cs.CV]. http://arxiv.org/abs/2110.05208

L. M. L. (n.d.). *Towards gender equity in Artificial Intelligence and machine learning applications in dermatology*. Journal of the American Medical Informatics Association : JAMIA. https://pubmed.ncbi.nlm.nih.gov/34151976/

Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S., & Lakshmanna, K. (2022). *Skin lesion classification of dermoscopic images using machine learning and convolutional neural network*. Scientific reports, 12(1), 18134. https://doi.org/10.1038/s41598-022-22644-9

Ward, A., Li, J., Wang, J., Lakshminarasimhan, S., Carrick, A., Campana, B., … Rao, P. (2024). Crowdsourcing dermatology images with Google Search ads: Creating a real-world skin condition dataset. Retrieved from http://arxiv.org/abs/2402.18545

Pacheco, A. G. C., Lima, G. R., Salomão, A. S., Krohling, B., Biral, I. P., de Angelo, G. G., Alves, F. C. R., Jr, Esgario, J. G. M., Simora, A. C., Castro, P. B. C., Rodrigues, F. B., Frasson, P. H. L., Krohling, R. A., Knidel, H., Santos, M. C. S., do Espírito Santo, R. B., Macedo, T. L. S. G., Canuto, T. R. P., & de Barros, L. F. S. (2020). PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data in Brief, 32(106221), 106221. https://doi.org/10.1016/j.dib.2020.106221