Utilizing Computer Vision, Clustering and Neural Networks for Complex Skin Disease Categorization

Ava Podrazhansky, Jennifer Nguyen, Rebekah Roepke
Department of Compute Science
Kennesaw State University
Marietta, GA, U.S.A.
{avapod, j15nguyenn, rebekah.roepke}@gmail.com

Dan Chia-Tien Lo
Department of Compute Science
Kennesaw State University
Marietta, GA, U.S.A
dlo2@kennesaw.edu

ABSTRACT

With the rise in premiums and deductibles for healthcare insurance, the availability for affordable medical care is becoming out of reach. Because of the diminished coverage, more people are opting to visit medical doctors only on an extreme need basis. Specifically, in visual dermal issues, like Melanoma, they are often neglected because of the high medical costs for consultations and the lack of public knowledge of symptoms. While some types of Melanoma could be benign, if the more malignant cases go undiagnosed it can lead to cancer and ultimately death. In this study, we propose a software approach to mitigate this problem, which utilizes artificial intelligence with high-level image processing to diagnose and categorize the most common forms of Melanoma. Our preliminary results indicate a viable approach with an accuracy higher than 85%.

CCS CONCEPTS

• Neural Networks • Machine Learning • Linear Regression •

KEYWORDS

Melanoma, Ridge Regression, Lasso Regression, K-Nearest Neighbor, Radial Basis Function

Introduction

Computer vision and machine learning have been active in research for several decades. The popularity of big data and the availability of high performance computing have brought artificial intelligent and machine learning to the spot light. A wide variety of data intensive and CPU-bound applications are becoming feasible recently. Several studies have been on Melanoma classification [1-4]. In this study, we propose to develop a system that integrates computer vision and neural networks to classify complex skin diseases. This system would benefit the general public to self-diagnose skin diseases in their early stages.

All images used for training in this study are from the international Skin Imaging Collaboration: Melanoma Project, and academic and industry partnered public image libraries. Figure 1 shows a sample image of such. A local dermatology clinic that specializes in dermatology will evaluate and verify the results. Additionally, the local dermatology clinic will be consulted for recommendations on the different types of Melanoma in terms of severity.

Artificial Intelligence Models

For the initial design of the model, several supervised machine learning models will be implemented. Specifically, two clustering algorithms such as K Nearest Neighbor (KNN) and Decision Trees (DT) and convolutional neural networks (CNN) with various activation functions will be studied and implemented. We will then use our results to design a modular deep learning neural network to accurately predict skin diseases.

We will start with a simple KNN model. Initially, the hyper-parameter k is set to 5, i.e., classification based on the 5 nearest neighbors using Euclidean distance. Several experiments with changing the value of k will be conducted and analyzed. Meanwhile, different distance measures will be studied such as Chi square distance and Minkowsky correlation distance. A subset of the attributes will be tested to study the relation of the performance with respect to the feature dimension.

A comparison of our KNN results with the Radial Basis Function (RBF) results. In this performance comparison, one of the foci is using all neighbors versus a selected number of neighbors.

A feed forward network will be adopted for further experiments. Different biases and activation functions will be studied. Several functions will be implement such as binary activation function, linear activation function, sigmoid function, and ReLU function. Some of these will be trained using backpropagation.

Utilizing Computer Vision, Clustering and Neural Networks for Complex Skin Disease Categorization

ACMSE '20, April, 2020, Tampa, Florida USA

Regression Types

To improve the performance of the Neural Networks, we have worked to find the most optimal types in regression testing for this topic. The main three to be implemented is Ridge, Lasso and Logistic Regression. With the usage of Ridge and Lasso regression, we can fit the model but decrease the risk of overfitting by minimizing or completely eradicating the impact of irrelevant features on the trained model. Because of the nature of the data being a combination of images and binary preliminary questions, we will be using Logistic Regression to accommodate those binary variables.

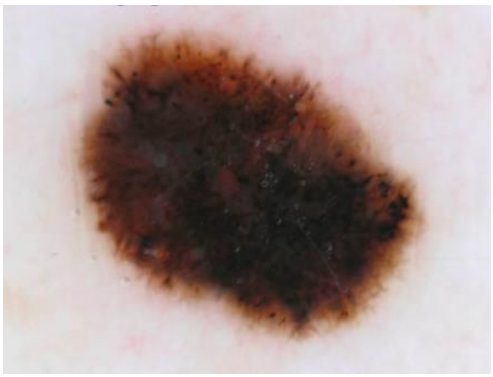

Figure 1: Malignant Melanoma

Decisions Trees

Decision trees will be used to reduce the noise to enhance the classification accuracy. We also plan on using Random Forest (RF) to classify Melanoma types from the images. The input will be passed down through numerous decision trees that represent each feature of Melanoma. This will allow the model to provide a higher accuracy of our diagnostics. For each decision tree, the input image will be compared to the categorized types of Melanoma from the database. The decision trees are important because they will identify the variables of the Melanoma types of an image.

Future Work

In the future, we hope to expand the usage of our model to utilize machine vision via a mobile application which would classify live photos rather than pre inserted set of images. Neural networks such as RNN can be used to monitor and compare the outcomes from using supervised learning. A prequestionnaire can also be included to obtain specific attributes, similar to the ones from our prebuilt dataset, (i.e. gender, age). We want to classify the images with unsupervised learning to compare the outcomes from using supervised learning. Additional to those functionalities, we would like to be able to enforce stricter image processing which would allow for measuring of the skin abnormality and use that to track growth and compare against regular growth rates.

CONCLUSION

Due to the lack of accessibility to medical care, people often wait until their case is extreme before consulting a medical professional. We want to decrease the number of severe skin disease cases by helping people identify skin diseases before the disease progresses into a critical state and monitor progress without requiring expensive, repeat consultations to a specialist. This will reduce severe skin disease cases by raising awareness of symptoms and providing diagnoses to people who would normally be unable to obtain professional help.

ACKNOWLEDGMENTS

The authors would like to thank reviewers for your valuable comments.

REFERENCES

[1] Matt Dennis, Judith Masthoff, Kirsten A. Smith, Peter Murchie, and Susan Hall. 2015. Designing a Tablet-based Intervention to Support Self-checking for Melanoma. In Proceedings of the 5th International Conference on Digital Health 2015 (DH '15). Association for Computing Machinery, New York, NY, USA, 79–86. DOI:https://doi.org/10.1145/2750511.2750518

[2] John David Osborne, Song Gao, Wei-bang Chen, Aleodor Andea, and Chengcui Zhang. 2011. Machine classification of melanoma and nevi from skin lesions. In Proceedings of the 2011 ACM Symposium on Applied Computing (SAC '11). Association for Computing Machinery, New York, NY, USA, 100–105. DOI:https://doi.org/10.1145/1982185.1982210

[3] Aaron N. Richter and Taghi M. Khoshgoftaar. 2018. Melanoma Risk Prediction with Structured Electronic Health Records. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB '18). Association for Computing Machinery, New York, NY, USA, 194–199. DOI:https://doi.org/10.1145/3233547.3233561

[4] Yunhao Ge, Bin Li, Yanzheng Zhao, Enguang Guan, and Weixin Yan. 2018. Melanoma Segmentation and Classification in Clinical Images Using Deep Learning. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing (ICMLC 2018). Association for Computing Machinery, New York, NY, USA, 252–256. DOI:https://doi.org/10.1145/3195106.3195164