

Development of Automated, Interactive Software for Skin Cancer Diagnosis using Image Analysis and Multivariate Modeling Techniques

In this project, an automated, interactive skin cancer diagnosis software was created to aid in early and reliable identification of skin cancer. This software is crucial for patients to be able to monitor their skin lesions for abnormalities between regular doctor visits, as taking action early allows cancerous lesions less time to develop. This software made multiple improvements on existing programs, including automatic processing of the lesion image, an engaging user interface, new analysis variables and programming methods, and a model with high diagnosis accuracy. The program was built in a more accessible programming language, JMP Scripting Language (similar to the commonly employed javascript), in order to address the inaccessibility of expensive software, such as MatLab. After 152 images of skin lesions were collected via online image search, multiple programming algorithms, including a new method for removing noise from the image, were created to isolate the outer border of the lesion from the original picture. Then, programs were created to analyze four categories of variables that determine the likelihood of a lesion's being cancerous: asymmetry, border irregularity, color variation, and texture, which was a newly introduced variable that was found to be very important. After these programs generated numerical values for each variable, a statistical model was created to classify the lesion as either harmless or cancerous. The model had a very low misclassification rate of 5.2%, and was validated again using a different method, demonstrating the effectiveness of the software. Several techniques developed in this project, such as novel methods for symmetry and texture analysis, made advances on previous research, and other techniques are well-known, reliable methods. There were great strides made in the skin cancer diagnostics field in this study, and future research will allow for even more program refinements and dataset increases.