Analytics for Hemoglobin Count Based on Image Data

Sam Mathew
Consultant
Gyan Data Private Limited
Chennai - 600 113 India
sam@gyandata.com

About the Client

En'Urga Inc. is an Indiana (US) based company leading in the cutting edge application of spray patternation technology and development of optical diagnostic equipments for single and multi-phase reacting flows. Located at the Purdue Research Park, En'Urga Inc. specializes in R&D and Proof of Concept (POC) applications as well as customization and calibration of optical diagnostic equipment. The application areas broadly include measurement of temperature, gas concentration, emissivity and particulate characteristics for a wide variety of industrial sectors and leading utilities.

Motivation

Data analysis has been practiced in medicine ever since its inception; for diagnosis and cure through clinical trials. Computer vision based approaches though have made progress into medicine only recently, especially for medical diagnostics. This is evident from the fact that 87.5% of the technical and research articles in Computer Vision-based Analytical Chemistry (CVAC) procedures have been published just in the last decade [1].

Hemoglobin count is essential for detecting anemia in short time. However, traditional routines of measurement of hemoglobin count is limited by the availability of safe, quick and efficient infrastructure required for phlebotomy

and lab testing which pose great challenges in times of epidemics. Hence, image based techniques like CVAC procedures are being explored by healthcare providers for Hemoglobin prediction from images of eye conjunctiva.

Problem

CVAC procedures are quite in their nascent conceptual stage and reasonable conclusions can only be derived through controlled experiments. The project envisioned building a statistical model for Hemoglobin prediction from images of eye conjunctiva taken on cattle in the field and in the sheep-pens.

The initial experiments were carried out by the research team operating out of Purdue. Over 100 images for sheep's palpebral conjunctiva were analyzed and the RGB (red-green-blue) channels from the region of interest were extracted. **Gyan Data** was tasked with building a predictive model for Hemoglobin count based on the RGB data and a host of derived parameters with little knowledge of the actual application due to the sensitivity of the project. During the course of interaction with the client, Gyandata was able to uncover a far valuable problem definition through persistent discussions and analysis of the domain knowledge. Our research revealed the necessity for controlled environment during image acquisition which was observed to be uncontrolled in the project.

Solution

The initial scope for Gyan Data's engagement was to apply different predictive models on just over 100 data points, with 10 variables in the input space and one predictor variable, the Hemoglobin count. Initial attempts focused on applying a host of model building techniques as simple as ordinary least-square (OLS) to more advanced kernel-Principal Component Regression (kPCR) with leave-one-out-cross-validation (LOOCV) for the small data-set. These efforts were preceded by data visualization and exploration that together confirmed the inadequacy of the naive regression methods.

Based on the variable meta data and through detailed investigation of the client's limited inputs, we could explore published research for similar applications. This revealed that during image acquisition, uneven lighting severely impacted the R, G, B channels of the image of palpebral conjunctiva. Consequently we suggested controlled lighting and a photographic standard for subsequent experiments.

Based on the approach followed in the published literature [2], an OLS model was built with derived features that were based on RGB channels in the conjunctiva image scaled by a photographic standard that was available in all images. In order to have uniform lighting though, this approach could only be applied on an even reduced set of data points. Recommendations were subsequently provided for future controlled experiments.

References

[1] Luis Fermín Capitán-Vallvey, Nuria López-Ruiz, Antonio Martínez-Olmos, Miguel M. Erenas, and Alberto J. Palma. Recent developments in computer vision-based analytical chemistry: A tutorial review. *Analytica Chimica Acta*, 899(Supplement C):23 – 56, 2015.

[2] Selim Suner, Gregory Crawford, John McMurdy, and Gregory Jay. Non-invasive determination of hemoglobin by digital photography of palpebral conjunctiva. *The Journal of emergency medicine*, 33(2):105–111, 2007.