**Skin texture change detection for clinical trials**

DISSERTATION

Submitted in partial fulfillment of the requirements of the

MTech, Data Science and Engineering

By

(NAGA KALYAN C S SARMA V)

(2018AH04042)

Under the supervision of

VENKATA KRISHNA TEJASVI SARRAJU

SENIOR ASSOCIATE,

COGNIZANT TECHNOLOGY SOLUTIONS

Dissertation work carried out at

COGNIZANT TECHNOLOGY SOLUTIONS,

HYDERABAD

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

Pilani (Rajasthan) INDIA

(November, 2020)

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**CERTIFICATE**

This is to certify that the Dissertation entitled “**Skin texture change detection for clinical trials”** and submitted by Mr. **NAGA KALYAN C S SARMA V** ID No: **2018AH04042** in partial fulfillment of the requirements of DSE CL ZG628T Dissertation embodies the work done by him under my supervision.

(Signature of the Supervisor)

**Place:** Hyderabad VENKATA KRISHNA TEJASVI SARRAJU

**Date:** 4th November, 2020 SENIOR ASSOCIATE

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**SECOND SEMESTER 2019-2020**

**DSECLZG628T DISSERTATION**

**Dissertation Outline**

**BITS ID No.**:2018AH04042

**Name of Student**: NAGA KALYAN C S SARMA V

**E-mail ID of the student**: 2018AH04042@wilp.bits-pilani.ac.in

**Name of Supervisor**: Venkata Krishna Tejasvi Sarraju

**Designation of Supervisor**: Senior Associate

**Qualification**: B.Tech **Experience**: 11 years

**E-mail ID of Supervisor**: TejasviVenkataKrishna.Sarraju@cognizant.com

**Title of Dissertation**: Skin texture change detection for clinical trials

**Name of First Examiner**: Prof. Sankara Nayaki K

**Designation of First Examiner**:

**Qualification and Examiner**:

**E-mail ID of First Examiner:**

**Name of Second Examiner**:

**Designation of Second Examiner**:

**Qualification and Examiner**:

**E-mail ID of Second Examiner:**

(Signature of Student) (Signature of Supervisor)

Date: 4th November, 2020 Date: 4th November, 2020

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**Work Integrated Learning Programmes Division**

**SECOND SEMESTER 2019-20**

**DSE CL ZG628T DISSERTATION**

**(EC-2 Mid-Semester Progress Evaluation Sheet)**

**NAME OF THE STUDENT:** NAGA KALYAN C S SARMA V

**ID NO. :** 2018AH04042

**Email Address :** 2018AH04042@wilp.bits-pilani.ac.in

**NAME OF SUPERVISOR :** VENKATA KRISHNA TEJASVI SARRAJU

**PROJECT TITLE :** Skin texture change detection for clinical trials

**EVALUATION DETAILS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **EC No.** | **Component** | **Weightage** | **Comments**  (Technical Quality, Originality, Approach, Progress, Business value) | **Marks Awarded** |
| 1 | Dissertation Outline | 10% |  |  |
| 2. | Mid-Sem Progress  Seminar  Viva  Work Progress | 10%  5%  15% |  |  |

|  |  |  |
| --- | --- | --- |
|  | **Supervisor** | **Additional Examiner** |
| **Name** | Venkata Krishna Tejasvi Sarraju | Prof. Sankara Nayaki K |
| **Qualification** | B.Tech |  |
| **Designation & Address** | Senior Associate,  Cognizant Technology Solutions,  Hyderabad |  |
| **Email Address** | TejasviVenkataKrishna.Sarraju@cognizant.com |  |
| **Signature** |  |  |
| **Date** |  |  |

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**Work Integrated Learning Programmes Division**

**SECOND SEMESTER 2019-20**

**Supervisor’s Rating of the Technical Quality of this Dissertation Outline**

EXCELLENT / GOOD / FAIR/ POOR (Please specify): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Supervisor’s suggestions and remarks about the outline (if applicable).**

Date: 4th November, 2020 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (Signature of Supervisor)

**Name of the supervisor:** VENKATA KRISHNA TEJASVI SARRAJU

**Email Id of Supervisor:** TejasviVenkataKrishna.Sarraju@cognizant.com

**Mob # of supervisor:** 9989787697

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**SECOND SEMESTER 2019-20**

DSE CL ZG628T **DISSERTATION**

Dissertation Title : SKIN TEXTURE CHANGE DETECTION FOR CLINICAL TRIALS

Name of Supervisor : TEJASVI VENKATA KRISHNA SARRAJU

Name of Student : NAGA KALYAN C S SARMA V

ID No. of Student : 2018AH04042

**Abstract**

In most of the clinical trials, specifically in cosmetic industry, a substantial amount of images are generated such that, the research scientists have an opportunity to analyze and document the efficacy of a chemical/product that is going to be introduced into the consumer market.

This poses 2 major problems:

1. The industries are bound by confidentiality agreements with the participants, hence each of these images are to be masked in order to protect PII from the internal and external technicians/approvers/stakeholders who view these images.
2. The research scientists are needed to analyze huge number of images that are generated over a period of time to understand good and bad effects of the product used in the trial, which is a cumbersome task and prone to human error

By leveraging Computer Vision algorithms in this solution, the facial parts of the participants which are not helpful for research (such as eyes) are automatically masked and also the skin texture changes can be compared from one image to another.

(Signature of Student) (Signature of Supervisor)

Date: 4th November, 2020 Date: 4th November, 2020

## Contents

1. Problem Statement: 8

2. Existing system: 8

3. Solution 8

4. Scope of work 8

5. Methodology 9

6. Uniqueness of the project 9

7. Benefit to the organization 9

8. Potential challenges & risks 9

9. Source 9

10. Input/Output Dimensions 9

11. Feasibility study 10

12. Resources needed 10

13. Plan of work 10

14. Solution architecture**:** 10

15. Work accomplished so far 11

15.1 Gather images with diversified demographics 11

15.2 Mask eyes in the facial images 12

15.3 Deploy to AWS 12

**Chapters**

**1. Problem Statement**:

The global skin care products market size was valued at 134.8 billion in 2018 and is projected to expand at a CAGR of 4.4% from 2019 to 2025 on account of the rapidly expanding global cosmetics industry. The growing e-commerce sector is anticipated to boost the market growth further. Moreover, technological and product innovations have led to an increased demand for skincare products.

This escalating demand for face creams, sunscreens and body lotions across the globe is pushing manufacturers to spend more time, money and resources on research and development. To target global markets, clinical trial participants are chosen from various demographics including age, location, ethnicity and skin types. Huge amounts of images get generated during the course of each trial, which are analyzed by the clinicians to note the efficacy of each chemical/product.

These industries are also bound by confidentiality agreements with the participants which demands protection of PII data.

**2. Existing system**:

To protect PII data, image masking softwares are used to cover some facial parts of the participants however, the licensing costs become a burden very soon. Moreover, manual observation of skin changes such as wrinkles, blemish etc by a clinician over such a large volume of images is time consuming and error prone.

**3. Solution**:

Build automated solution to mask facial parts (which need not be analyzed) for protecting identity of the participants. Use computer vision algorithms to detect changes in skin texture and provide insights to the clinicians to better understand the efficacy of the products.

**4. Scope of work:**

The scope of work will involve understanding the design and other requirements

* Explore the machine learning and deep learning techniques for image masking
* Experiment with various Computer Vision algorithms for texture analysis
* Design the solution in cloud to meet the algorithm and visualization requirements

**5. Methodology**:

1. Clinicians to upload the raw images to cloud storage
2. Mask the images instantly and delete the original images
3. Apply computer vision algorithm to perform texture analysis on facial images
4. Build visualizations for the clinicians to better understand the efficacy of the products

**6. Uniqueness of the project:**

This is an end-to-end solution rather just a portion of what is expected by the industry. The following modules form the entire solution:

* Data engineering: Gathering, storing and masking of data
* Data Science: Applying Deep Learning and Computer Vision algorithms to generate insights from image which are helpful for research
* Operationalization: Leveraging cloud services to host the entire solution in order to generate repeated value

**7. Benefit to the organization:**

The following are the primary benefits:

* Ability to automatically infer the changes in the skin texture, thereby identifying efficacy of each chemical/product being used in the lab
* Compliance with confidentiality agreements with the participants of the clinical trials
* Since no human involvement is needed, scalability in terms of speed, volume and accuracy is a huge benefit.

**8. Potential challenges & risks:**

* Gathering huge volumes of images to train the model
* The masking process has to be 100% accurate to stay compliant
* Good amount of testing is needed in skin texture identification as key decisions are taken around the chemicals used in a product

**9. Source:**

Since we are not allowed to use the actual images taken in the lab, in order to train/test the algorithm and for user acceptance testing purposes, we will be downloading free image datasets available on ImageNet, Kaggle, Google Images, celebfaces etc to gather a collection of facial images suitable for our use case.

**10. Input/Output Dimensions :**

Inputs will be ID, Ethnicity, timing\_of\_photograph, age, gender and product\_under\_testing provided as S3 tags associated with facial images of the participants. Output 1 will be Masked images and Output 2 will be histogram for each "region" of an input facial image (no. of regions can be controlled by the user to get higher or lower granular details) to help analyze skin texture (such as wrinkles). Other visualizations according to Ethnicity, age group are being planned.

**11. Feasibility study:**

Although lot of aspects of this project are experimental in nature, the planned outcome is certainly possible to achieve as long as we take a controlled and step-by-step approach for the completion of the project.

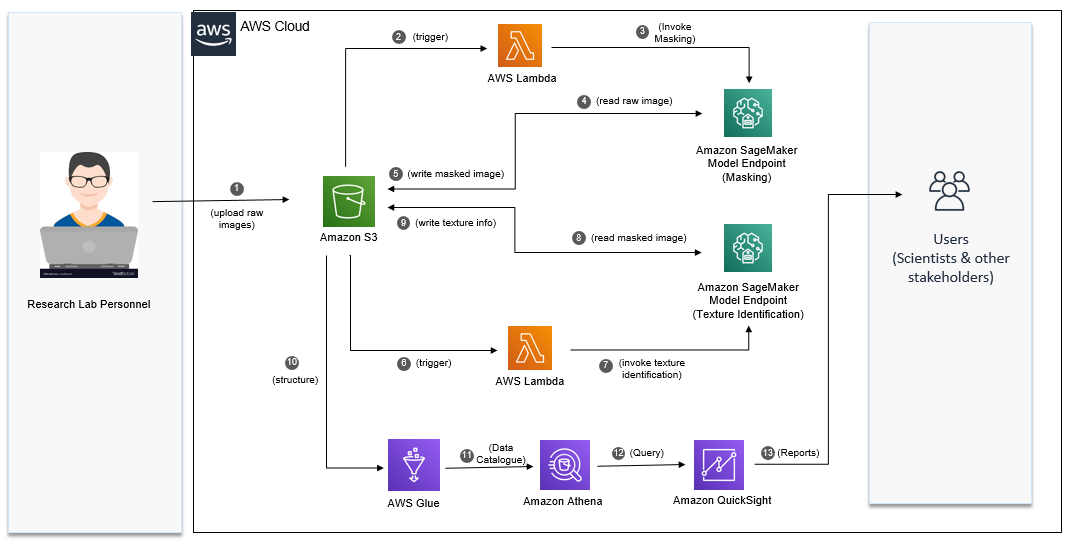
**12. Resources needed**:

Cloud platform, Python, Computer Vision, Machine Learning and Deep Learning packages and good number of images for training/testing

**13. Plan of work:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No.** | **Task** | **Planned duration (in weeks)** | **Name of Deliverables** |
| 1 | Research and feasibility check | 1 | Research |
| 2 | Gathering the raw images | 1 | Data Gathering |
| 3 | Setup a cloud platform with storage, compute and software | 1 | Cloud setup |
| 4 | Implement PII data masking | 5 | Application |
| 5 | Implement texture analysis using a computer vision algorithm | 5 | Application |
| 6 | Build visualization showing efficacy of products | 3 | Reporting |

**14. Solution architecture:**



**15. Work accomplished so far:**

15.1 Gather images with diversified demographics

Gathered images that concur to the following criteria:

* Clear image with high pixel density and light background
* Subject (the participant) in the image should be looking into the camera
* Face of the subject should be entirely present in the picture
* Subjects should belong to diversified demographics such as Age, Ethnicity, Gender

15.2 Mask eyes in the facial images

Leveraged OpenCV library in Python to detect eyes in the facial images and mask them.

15.3 Deploy to AWS

The images have been uploaded to a S3 bucket. A lambda function would be polling for incoming images. As soon as an image arrives, OpenCV code is executed and masked image is stored as an output.