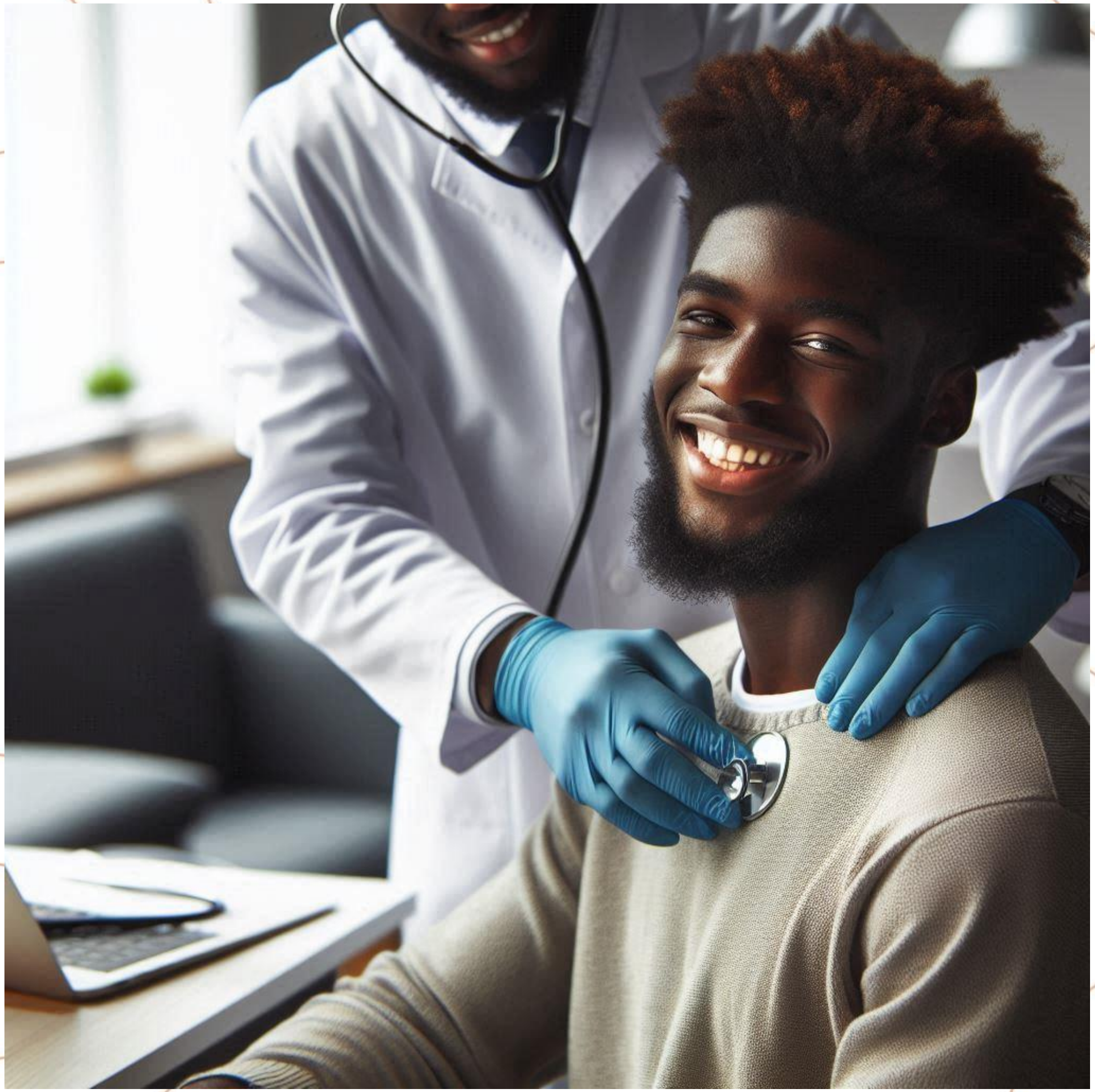
10Alytics

SQL CAPSTONE PROJECT

Analysis of DermAI Diagnostics Skin Cancer Dataset

10Alytics



Business Introduction

Skin cancer is one of the most common and life-threatening diseases, yet early detection can significantly improve survival rates. Our business, DermAI Diagnostics, leverages machine learning and clinical dermatology research to enhance early diagnosis and treatment. By analyzing patient demographics, environmental factors, and lesion characteristics, we create data-driven insights to assist dermatologists in early-stage detection and decision-making.

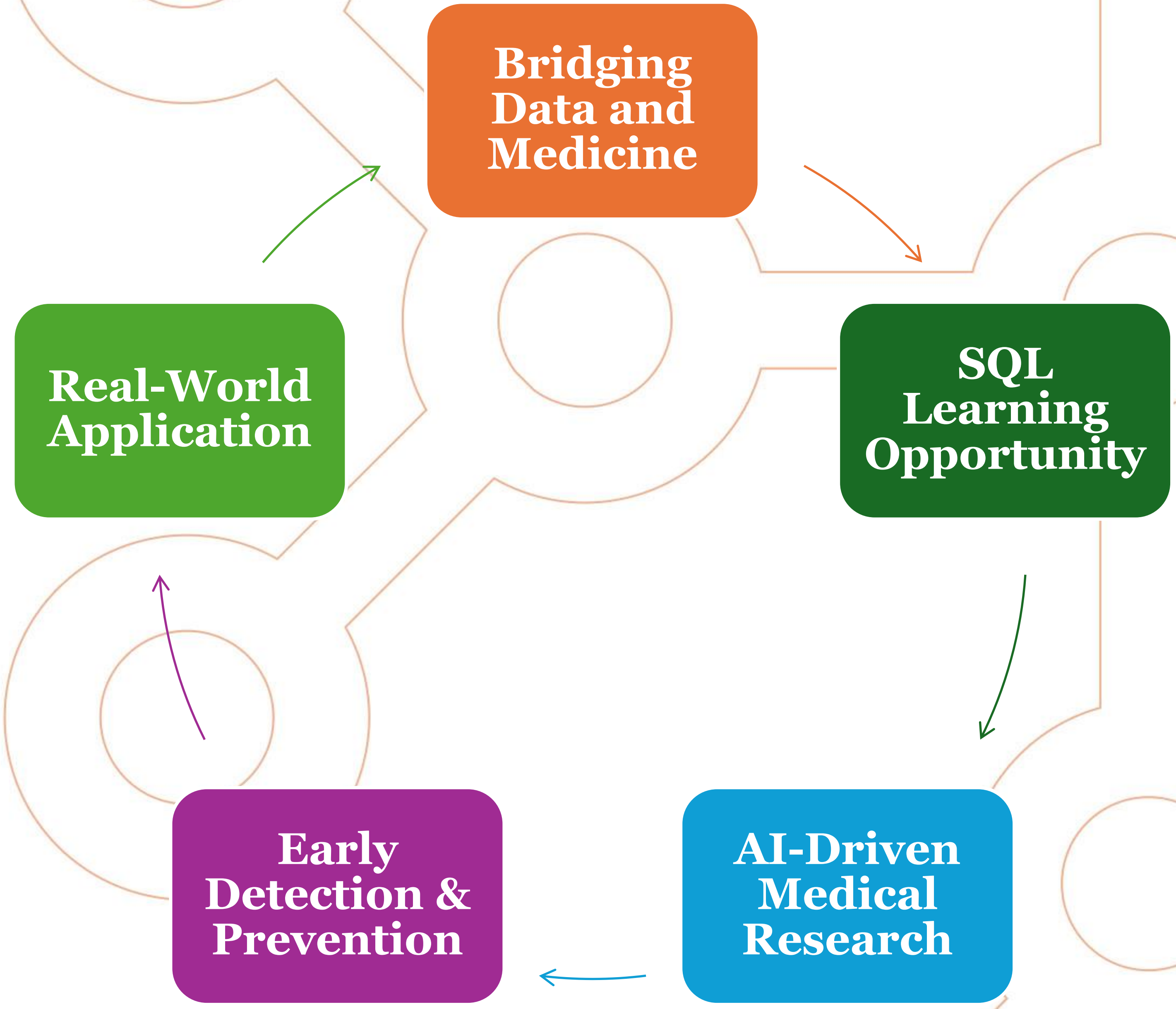
Through a combination of AI-powered diagnostic tools, real-world clinical data, and SQL-based research, our company aims to bridge the gap between medical practitioners and machine learning-based skin lesion classification. By providing a structured dataset and digital tools, we support medical research, epidemiological studies, and AI-driven skin cancer detection, ultimately improving public health outcomes.

PROBLEM STATEMENT

Skin cancer detection is often delayed due to misdiagnosis, lack of access to dermatologists, and limited understanding of environmental risk factors. With 1,089 instances of skin lesions in our dataset, we aim to uncover key patterns that link demographics, environmental exposure, and lesion characteristics to different types of skin cancer. This project seeks to enhance early-stage diagnosis and machine learning-based decision support by structuring the data for SQL queries, analysis, and model training.

10Alytics

Rationale for the Project



Bridging Data and Medicine

SQL Learning Opportunity

AI-Driven Medical Research

Early Detection & Prevention

Real-World Application

10Alytics

DATA DICTIONARY

Table 1: Patient_Info

Column Name	Description
patient_id	Unique identifier for each patient
smoke	Patient smokes (TRUE/FALSE)
drink	Patient drinks alcohol (TRUE/FALSE)
background_father	Patient’s paternal ethnicity
background_mother	Patient’s maternal ethnicity
age	Age of patient
pesticide	Exposure to pesticides (TRUE/FALSE)
gender	Gender (MALE/FEMALE)
skin_cancer_history	Previous skin cancer diagnosis (TRUE/FALSE)
cancer_history	Family history of cancer (TRUE/FALSE)
has_piped_water	Access to piped water (TRUE/FALSE)
has_sewage_system	Access to sewage system (TRUE/FALSE)

10Alytics

DATA DICTIONARY

Table 2: Lesion_Info

Column Name	Description
lesion_id	Unique identifier for each lesion
patient_id	Foreign key linking to **Patient_Info**
fitspatrick	Fitzpatrick skin type (1-6)
region	Body region of the lesion
diameter_1	Diameter of lesion (mm)
diameter_2	Second diameter measurement (mm)
diagnostic	Type of skin lesion (BCC, MEL, NEV, etc.)
itch	Lesion causes itching (TRUE/FALSE)
grew	Lesion has grown (TRUE/FALSE)
hurt	Lesion causes pain (TRUE/FALSE)
changed	Lesion changed in color/size (TRUE/FALSE)

10Alytics

DATA DICTIONARY

Table 2: Lesion_Info

Column Name	Description
bleed	Lesion bleeds (TRUE/FALSE)
elevation	Lesion is raised (TRUE/FALSE)
img_id	Associated lesion image filename
biopsed	Whether the lesion was biopsy-confirmed (TRUE/FALSE)

10Alytics

--Aim of the Project—

Develop a SQL database for students to practice joining clinical and lesion data for effective skin cancer analysis.

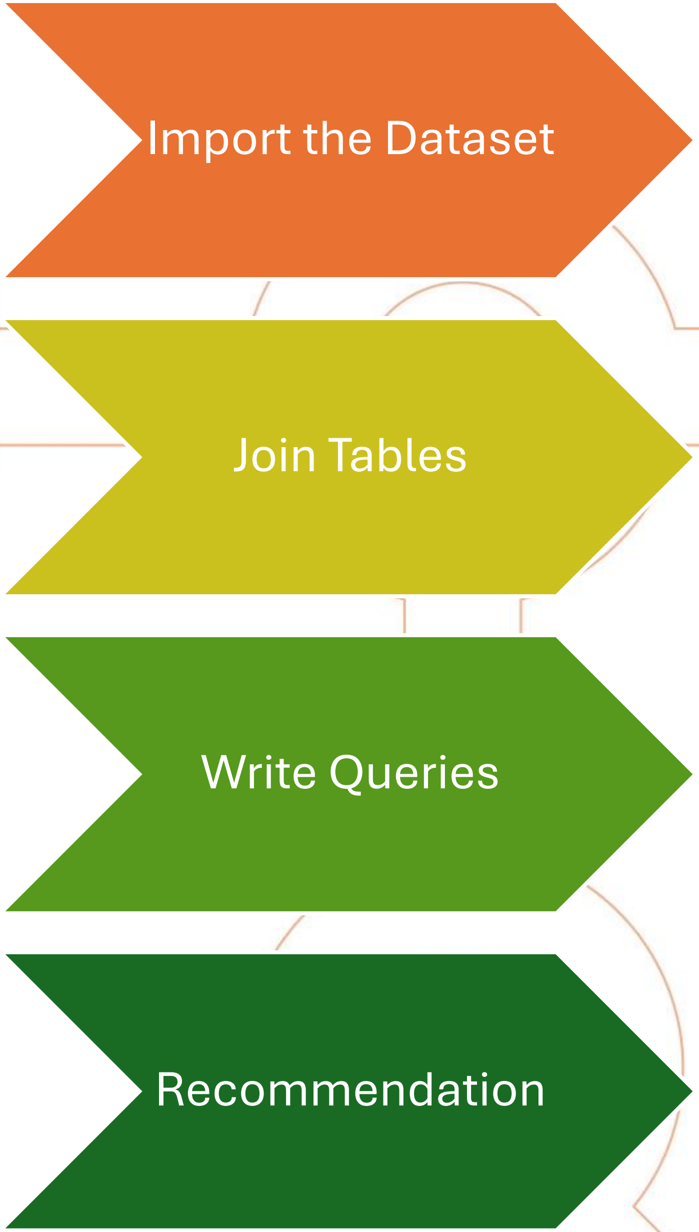
Identify environmental and demographic risk factors that correlate with specific skin lesions.

Analyze lesion characteristics to find patterns that indicate cancerous vs. benign lesions.

Create a machine learning-ready dataset that supports AI-based early detection of skin cancer using structured metadata.

Enhance dermatological research by providing a well-organized dataset for epidemiological studies and AI model training.

Project Workflow



Import the Dataset

Join Tables

Write Queries

Recommendation