



DFS - HealthCare Knowledge Graph

Team Dr. Data

(Sarthak Bansal, Krishna Singh, Deoptansh Sharma)

Functional Requirements

1. Data Ingestion:

- The system should be able to ingest data from various sources such as Oral Cancer, Heart Disease, Patient Database, and Diabetes Database.

2. Data Integration:

- Establish relationships between entities from different datasets to create a comprehensive knowledge graph.
- Define rules for merging and linking entities to ensure data consistency and accuracy.

3. Graph Construction:

- Develop Queries to construct a meaningful graph structure from the ingested data.
- Ensure that the graph accurately represents the conceptual relationships between medical entities from incoherent data acquired from different sources.

4. Query Language Support:

- Utilise Cypher query language effectively to retrieve information from the knowledge graph.
- Support complex queries to enable detailed exploration of medical relationships and data.

5. Scalability:

- Design the system to handle a growing volume of medical data.
- Ensure that the knowledge graph can scale horizontally to accommodate increasing data and user demands.

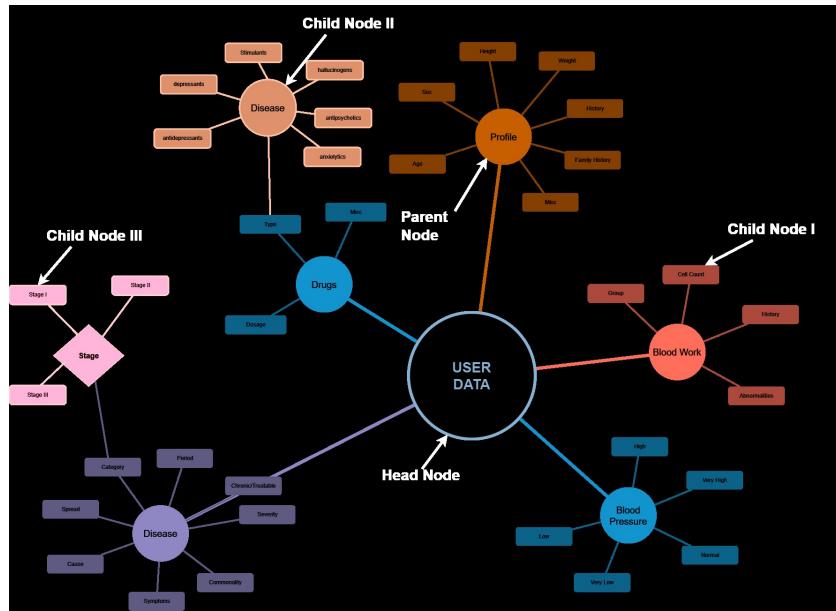
6. User Interface:

- Develop an intuitive user interface that allows users to interact with the knowledge graph easily.
- Provide visualization tools to represent the graph and facilitate exploration.

7. Search and Retrieval:

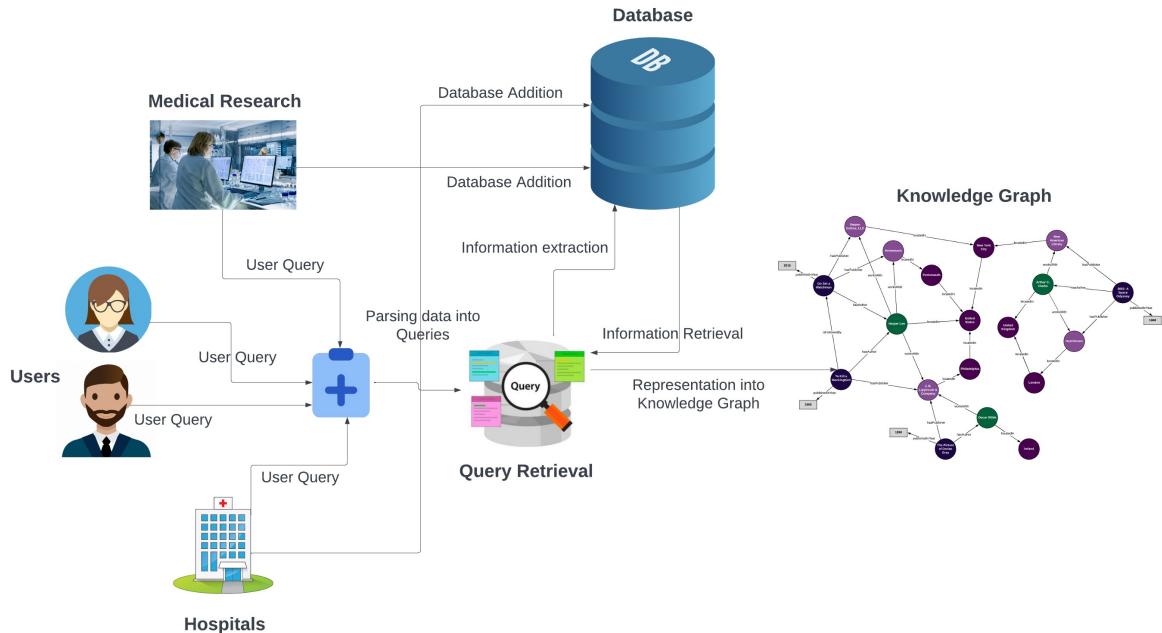
- Implement a robust search mechanism to quickly retrieve relevant information from the knowledge graph.

What is a Knowledge Graph?



A knowledge graph is a way of organizing and representing information, allowing for connections and relationships between different pieces of data. It's a graph structure that typically consists of nodes (entities) and edges (relationships) that define how these entities are interconnected. In healthcare domain, our knowledge graph represents nodes, which include entities like (Disease, Blood Pressure, Blood Work, Medicines Prescribed, Stage of Disease, User Profile etc..). The nodes are further subdivided into head, child and leaf nodes based on their importance/relevance to the data. The nodes are connected via edges that define relationships between nodes created from incoherent data.

Application/Use Case



This system is designed to be used by individuals, hospitals, or organizations interested in the Healthcare domain.

- The users can view conceptual relationships between data acquired from different sources.
- Users interact with the form to provide details to the query, which is then retrieved from the database to generate the Knowledge Graph.
- The primary users also act as a source to add data to the database.
- The Knowledge Graph acts as a visual aid for the users to look out for insights related to their queries.
- The users can also look into the data generated query graphs without sharing their information.

Non-Functional Requirements

1. Performance:

- *Response Time*: The system will respond to user queries within seconds.
- *Scalability*: The system when deployed is robust enough to handle multiple queries at once.

2. Reliability:

- *Data Integrity*: The data is taken from the centralized database. The data taken from the user as a query is without a name. Complete anonymization can be added in further release.

3. Scalability:

- *Horizontal and Vertical Scalability*: The system can be scaled to add more type of system queries as the data is scaled and improved.

4. Usability:

- *User Interface*: The user-face is self-explanatory and user friendly.
- Inferencing - Users can view conceptual relationships between data

5. Compatibility:

- *Browser and Device Compatibility*: The system is compatible with any browser and requires no plugins.
- *Integration*: It is easily integrable with the existing IHUB Data-Foundation's ecosystem.

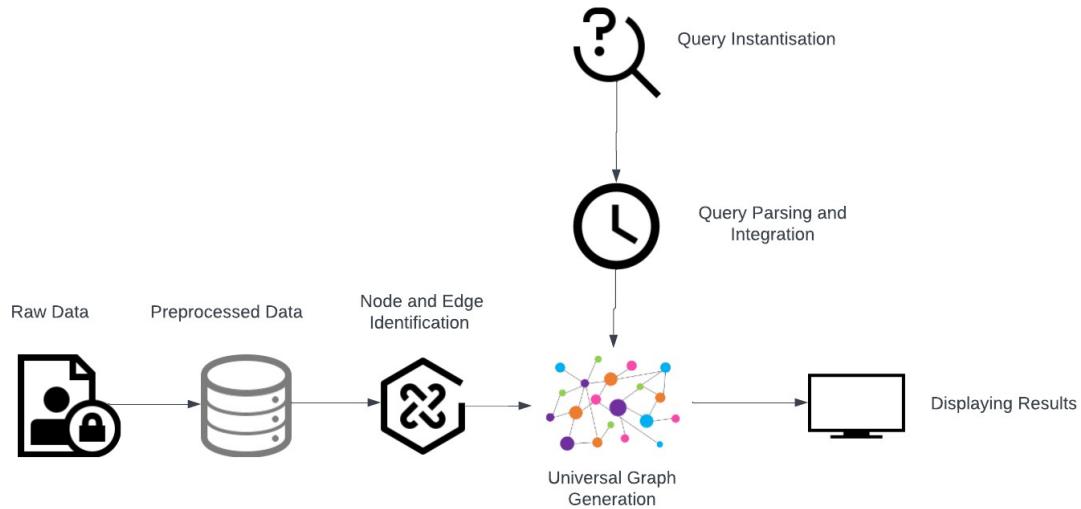
6. Maintainability:

- *Modifiability*: The system architecture allows for easy modifications and updates.

Stakeholders

The stakeholders include healthcare researchers, Medical industry, Self-Practitioners or any general inquisitive individual.

Design



- The raw data is cleaned and preprocessed into standard database headings.
- The relevant nodes and edges are identified and assigned.
- A universal graph is generated with all of the data.
- Multiple queries are written based on user input or from standard preset query parameters.
- The query is parsed and integrated into the existing Knowledge Graph.
- The final knowledge graph based on user query is displayed.

Testing

The system is robust and stable to external testing. We have added constraints to the input formats to prevent the users from uploading misclassified and vague data. This ensures that the data is as per our database standards. We also deployed testing scripts to test the vulnerabilities of our system.

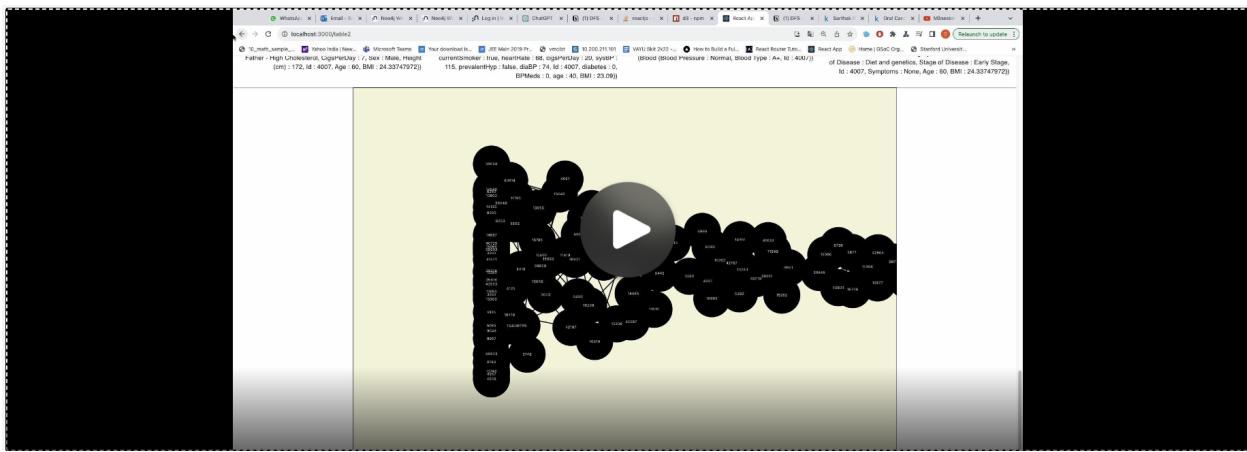
Histopathological Support

The system aims to provide histopathological support in future releases. The current database has entries labelled as Cancer and Non-Cancer based on EfficientNet Model which predicts the same.

Future release aims to extend this support for user uploaded images too.

Application Demo Results

Demo Video



https://iiitaphyd-my.sharepoint.com/:v/g/personal/sarthak_bansal_students_iit_ac_in/EfG_k3TXW5RdkHlpAY4YDWgBZl-NDkGfh0wHuePxLxNdQ?nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAiOjPbmVFcmI2ZUZyckJ1c2luZXNzIiwickmVmZXJyYWxBcHBQbGF0Zm9ybSI6lldlYiIsInJlZmVycmFsTW9

Database Files

	A	B	C	D
1	Id	Cancer	sysBP	prevalentHyp
2	0	0	146.786223	1
3	1	0	80.2890087	0
4	2	1	146.830428	1
5	3	1	149.29173	1
6	4	1	169.999895	1
7	5	1	163.990567	0
8	6	0	140.714256	1
9	7	0	132.412092	0
10	8	1	148.716318	1
11	9	0	130.289487	0
12	10	1	160.931634	0
13	11	0	130.192737	1
14	12	1	151.795229	1
15	13	0	144.232643	0
16	14	1	151.766729	0

Cancer Data

A	B	C	D	E	F	G	H	I	J	K	
1	Id	Gender	age	education	currentSmok	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol
2	1	Male	39	postgraduate	0	0	0	0	0	0	195
3	2	Female	46	primaryschol	0	0	0	0	0	0	250
4	3	Male	48	uneducated	1	20	0 no	0	0	0	245
5	4	Female	61	graduate	1	30	0 no	1	0	0	225
6	5	Female	46	graduate	1	23	0 no	0	0	0	285
7	6	Female	43	primaryschol	0	0	0 no	1	0	0	228
8	7	Female	63	uneducated	0	0	0 no	0	0	0	205
9	8	Female	45	primaryschol	1	20	0 no	0	0	0	313
10	9	Male	52	uneducated	0	0	0 no	1	0	0	260
11	10	Male	43	uneducated	1	30	0 no	1	0	0	225
12	11	Female	50	uneducated	0	0	0 no	0	0	0	254
13	12	Female	43	primaryschol	0	0	0 no	0	0	0	247
14	13	Male	46	uneducated	1	15	0 no	1	0	0	294
15	14	Male	41	graduate	0	0	1 no	1	0	0	332
16	15	Female	39	primaryschol	1	9	0 no	0	0	0	226
17	16	Female	38	primaryschol	1	20	0 no	1	0	0	221
18	17	Male	48	graduate	1	10	0 no	1	0	0	232
19	18	Female	46	primaryschol	1	20	0 no	0	0	0	291
20	19	Female	38	primaryschol	1	5	0 no	0	0	0	195
21	20	Male	41	primaryschol	0	0	0 no	0	0	0	195
22	21	Female	42	primaryschol	1	30	0 no	0	0	0	190

Heart Data

Node and Edge Identification

Database Information

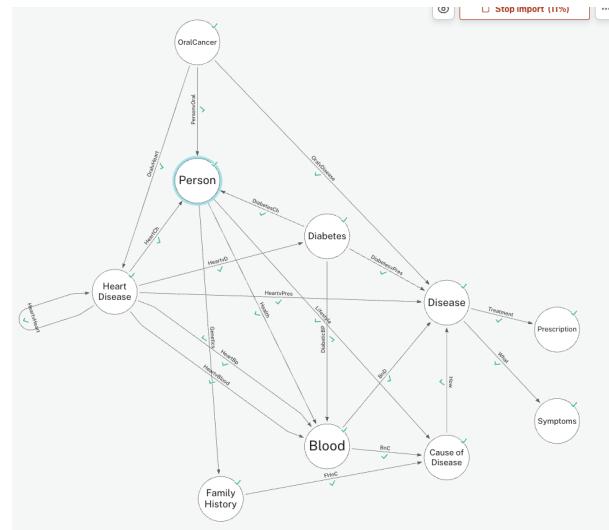
Nodes (121,280)



Relationships (204,566)

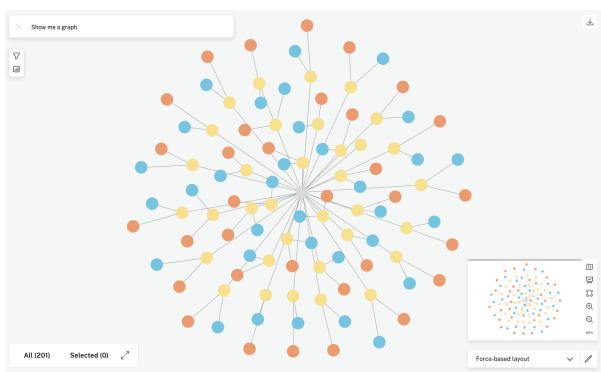


Node and Edge List

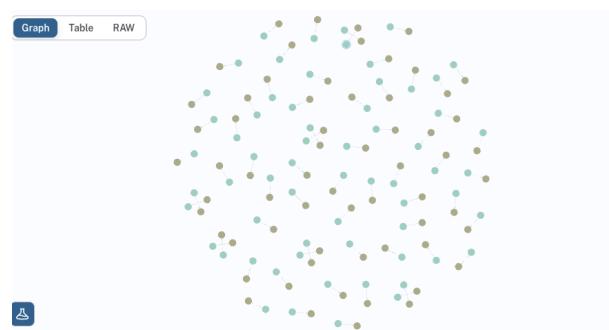


Relationship Map

Knowledge Graphs



Universal Knowledge Graph

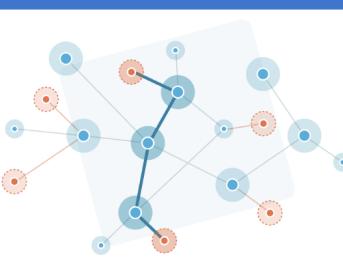


Query Graph on Neo4j

User Interface

Knowledge Graphs

A knowledge graph is a dynamic system that organizes information using graph structures. Entities and their relationships are represented as nodes and edges, facilitating advanced data analysis and automated reasoning in various domains.



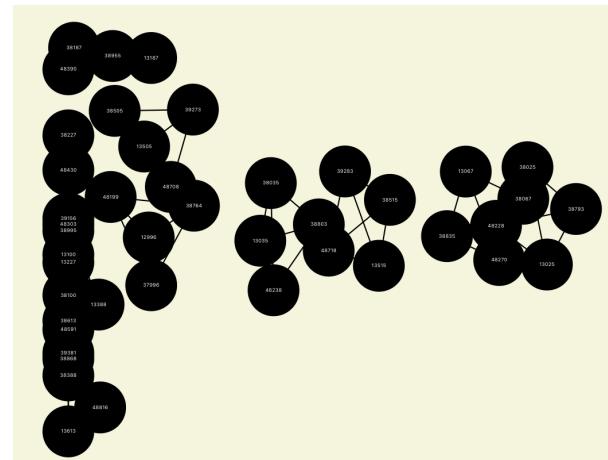
<h3>Preset Diabetes</h3> <p>Explore the conceptual connection between the general medical chart , general heart stroke data and general diabetic data</p>	<h3>Preset Heart Disease</h3> <p>Explore the conceptual connection between the general medical chart , general heart stroke data and general diseases and prescription data</p>	<h3>Preset Oral Cancer</h3> <p>Explore the conceptual connection between the general heart stroke , diabetic data ,oral cancer and diseases & prescription data</p>
GET STARTED	GET STARTED	GET STARTED

User Interface

Preset Query Results

Patient Demographic		Dolentes Data		Blood Pressure		Heart Disease	
Patient (Weight (kg): 64, Gender: Male, Family History: None, Cophydrase: 0, Sex: Male, Height (cm): 170, Age: 31, BMI: 24.28(2544))		(Dolentes) (Database) (DolentesFunction: -0.107, BloodPressure: 75, Outcome: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 27, Prostagrenes: 0, BM: 23.8)		(Blood Pressure) (Normal, Blood Type: Ad, Iu: 72)		(Heart Disease) (0, PRESENCE: 0, education: primarySchool: 12, HighSchool: 25, Master: 3, Doctorate: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 27, Prostagrenes: 0, BM: 23.8)	
Patient (Weight (kg): 70, Family History: Father: High Cholesterol, Brother: No, Cophydrase: 0, Sex: Female, Height (cm): 175, Age: 36, BMI: 24.61(2508))		(Dolentes) (Database) (DolentesFunction: -0.140, BloodPressure: 76, Outcome: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 21, Prostagrenes: 0, BM: 24.6)		(Blood Pressure) (Normal, Blood Type: Ad, Iu: 34)		(Heart Disease) (0, PRESENCE: 0, education: uneducated: 10, HighSchool: 21, Master: 0, Doctorate: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 21, Prostagrenes: 0, BM: 24.6)	
Patient (Weight (kg): 70, Family History: Father: High Cholesterol, Brother: Yes, Cophydrase: 0, Sex: Female, Height (cm): 175, Age: 36, BMI: 24.61(2508))		(Dolentes) (Database) (DolentesFunction: -0.143, BloodPressure: 76, Outcome: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 20, Prostagrenes: 0, BM: 24.6)		(Blood Pressure) (Normal, Blood Type: Ad, Iu: 36)		(Heart Disease) (0, PRESENCE: 0, education: uneducated: 10, HighSchool: 21, Master: 0, Doctorate: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 20, Prostagrenes: 0, BM: 24.6)	
Patient (Weight (kg): 70, Gender: Male, Family History: None, Cophydrase: 0, Sex: Male, Height (cm): 176, Age: 31, BMI: 33.88(201877))		(Dolentes) (Database) (DolentesFunction: -0.157, BloodPressure: 75, Outcome: 0, n: 264, Disease: 0, Shit/Stroke: 32, Inulin: 0-4, Age: 20, Prostagrenes: 2, BM: 21)		(Blood Pressure) (Normal, Blood Type: Ad, Iu: 37)		(Heart Disease) (0, PRESENCE: 0, education: primarySchool: 12, HighSchool: 25, Master: 3, Doctorate: 0, n: 264, Disease: 0, Shit/Stroke: 32, Inulin: 0-4, Age: 20, Prostagrenes: 2, BM: 21)	
Patient (Weight (kg): 70, Gender: Male, Family History: None, Cophydrase: 0, Sex: Male, Height (cm): 176, Age: 31, BMI: 33.88(201877))		(Dolentes) (Database) (DolentesFunction: -0.157, BloodPressure: 75, Outcome: 0, n: 264, Disease: 0, Shit/Stroke: 32, Inulin: 0-4, Age: 20, Prostagrenes: 2, BM: 21)		(Blood Pressure) (Normal, Blood Type: Ad, Iu: 17)		(Heart Disease) (0, PRESENCE: 0, education: primarySchool: 12, HighSchool: 25, Master: 3, Doctorate: 0, n: 264, Disease: 0, Shit/Stroke: 32, Inulin: 0-4, Age: 20, Prostagrenes: 2, BM: 21)	
Patient (Weight (kg): 75, Gender: Male, Family History: None, Cophydrase: 0, Sex: Male, Height (cm): 177, Age: 31, BMI: 23.71(2520))		(Dolentes) (Database) (DolentesFunction: -0.66, BloodPressure: 65, Outcome: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 21, Prostagrenes: 0, BM: 24.6)		(Blood Pressure) (Normal, Normal Blood Type: Ad, Iu: 33)		(Heart Disease) (0, PRESENCE: 0, education: primarySchool: 12, HighSchool: 25, Master: 3, Doctorate: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 21, Prostagrenes: 0, BM: 24.6)	
Patient (Weight (kg): 75, Gender: Male, Family History: None, Cophydrase: 0, Sex: Male, Height (cm): 177, Age: 31, BMI: 23.71(2520))		(Dolentes) (Database) (DolentesFunction: -0.66, BloodPressure: 65, Outcome: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 21, Prostagrenes: 0, BM: 24.6)		(Blood Pressure) (Normal, Normal Blood Type: Ad, Iu: 33)		(Heart Disease) (0, PRESENCE: 0, education: primarySchool: 12, HighSchool: 25, Master: 3, Doctorate: 0, n: 264, Disease: 0, Shit/Stroke: 14, Inulin: 0-4, Age: 21, Prostagrenes: 0, BM: 24.6)	

Tabular View



Graphical View

User Input for Data

My Form

Male Female

Diabetes in Family

SUBMIT

My Form

Male Female

Smoker

SUBMIT

User Form 1

User Form 2

User Query Data

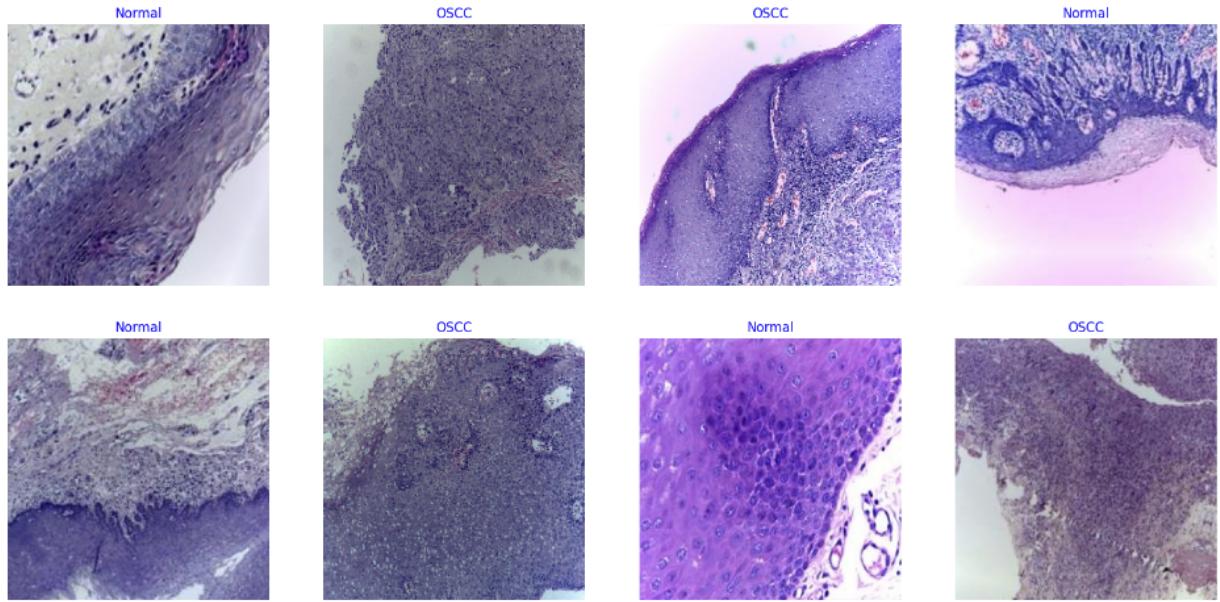
Person	Diabetic	Disease
(Person (Weight (kg) : 70, Family History : None, CigsPerDay : 4, Sex : Female, Height (cm) : 166, Id : 702, Age : 41, BMI : 26.21882086))	(Diabetes (DiabetesPedigreeFunction : 0.565, BloodPressure : 78, Outcome : 1, Id : 702, Glucose : 125, SkinThickness : 31, Insulin : 0, Age : 49, Pregnancies : 6, BMI : 27.48))	(Disease (Medicines Prescribed : Propranolol, Disease : Migraine, Sex : Female, Category of Medicine : Oral Hypoglycemic Agent, Cause of Disease : Genetic predisposition, Stage of Disease : Early Stage, Id : 460, Symptoms : Frequent urination, Age : 67, BMI : 27.00913077))

Tabular and Graphical View of User Query

Future Release

The future release aims to extend the support for EfficientNet Model for Cancer Detection to user based queries.

Sample Training Images



Model Architecture

Layer (type)	Output Shape	Param #
<hr/>		
efficientnetb3 (Functional)	(None, 1536)	10783535
<hr/>		
batch_normalization (BatchN ormalization)	(None, 1536)	6144
<hr/>		
dense (Dense)	(None, 256)	393472
<hr/>		
dropout (Dropout)	(None, 256)	0
<hr/>		
dense_1 (Dense)	(None, 2)	514
<hr/>		
Total params: 11,183,665		
Trainable params: 11,093,290		
Non-trainable params: 90,375		
<hr/>		

Training Results



Report and Matrix

	precision	recall	f1-score	support
Normal	0.99	0.99	0.99	360
OSCC	0.99	0.99	0.99	419
accuracy			0.99	779
macro avg	0.99	0.99	0.99	779
weighted avg	0.99	0.99	0.99	779

