Project Team 4

Conversational Agent for Disease Information Retrieval

Team members: Chintan Sheshagiri Addoni, Manasa Krishnan

Introduction

In the realm of healthcare, the ability to efficiently and accurately access information about rare diseases is crucial, not only for medical professionals but also for patients and their families. Our project aims to bridge the gap in information accessibility by developing a state-of-the-art conversational agent specifically designed to handle queries related to rare diseases. This innovative tool is built to navigate through extensive medical databases, utilizing advanced information retrieval models to ensure the accuracy and relevance of the information provided.

The core functionality of our conversational agent revolves around its sophisticated capability to interpret user inputs. By identifying key information about rare diseases, the agent can effectively address the specific needs and concerns of its users. This is achieved through the implementation of advanced semantic analysis techniques, which enable the system to comprehend the context of inquiries and provide a brief answer to the query.

To bolster this functionality, our team has undertaken the meticulous task of creating a detailed rare disease dataset. This dataset is being compiled through comprehensive data parsing and scraping techniques, forming the backbone of our information retrieval process. This process is strategically designed to rank the relevance of the information presented, ensuring that users are provided with the most pertinent and helpful details.

One of the distinguishing features of our conversational agent is its ability to present ranked documents in a summarized format. This approach significantly improves user readability and comprehension, which is particularly vital in navigating the complexities of medical information. The ultimate goal of our project is to empower users, enabling them to refine their searches and acquire the most relevant information efficiently. By doing so, we aim to contribute significantly to the field of medical information retrieval, particularly in the context of rare diseases, where accurate information can be pivotal in decision-making and understanding.

Task Description

1. Dataset Generation

The foundation of our project is the creation of a comprehensive dataset focused on rare diseases, a crucial resource for our conversational agent. We identified **Orphanet** and **PubMed** as key sources for high-quality research articles on rare diseases. Our dataset primarily encompasses rare respiratory, renal, teratologic, allergy, renal, ophthalmic, infertility, immunological, hepatic, and hematological diseases.

Key features of our dataset include:
Disease Information: Each disease is cataloged with a unique name and code.
Research Articles: Associated research articles are included, featuring details like article ID, title, abstract, link for the article, and full text.

Advanced web scraping techniques were employed to amass this data, resulting in a final dataset comprising 27,507 entries. This task was instrumental in laying the groundwork for our project, reflecting our commitment to achieving a grade **B level** of accomplishment.

Our dataset contains information related to the below categories of rare dieseases:

Disease	**Category**
Rare abdominal surgical diseases	Abdominal
Rare allergic diseases	Allergic
Rare disorder without a determined diagnosis after full investigation	Undetermined
Rare gastroenterological diseases	Gastrointestinal
Rare genetic diseases	Genetic
Rare haematological diseases	Haematological
Rare hepatic diseases	Hepatic
Rare immunological diseases	Immunological
Rare inborn errors of metabolism	Metabolic
Rare infertility	Reproductive
Rare odontological diseases	Dental

Rare ophthalmic diseases	Ophthalmic
Rare otorhinolaryngological diseases	ENT
Rare renal diseases	Renal
Rare respiratory diseases	Respiratory
Rare skin diseases	Skin
Rare surgical maxillo-facial diseases	Maxillofacial
Rare surgical thoracic diseases	Thoracic
Rare systemic and rheumatological diseases of childhood	Childhood
Rare systemic and rheumatological diseases	Systemic
Rare teratologic disorders	Teratologic
Rare urogenital diseases	Urogenital

2. Data Preprocessing

Our data preprocessing phase is a critical step in refining our dataset to bolster the efficiency and accuracy of our information retrieval system. We initiated this process with Text Simplification, involving the removal of citations, references, and stop words from the full texts of the articles. This step is crucial for eliminating irrelevant or redundant data, allowing for a more focused analysis of the content.

Subsequently, we employed Vectorization, utilizing the TF-IDF (Term Frequency-Inverse Document Frequency) technique. TF-IDF is a statistical measure that evaluates the importance of a word in a document, relative to a collection of documents. It helps in distinguishing the significance of words by considering their frequency across the dataset, thereby enhancing the relevance of our search results. By assigning higher weights to terms that are frequent in a document but rare across the dataset, TF-IDF enables our model to identify and prioritize key information effectively.

These preprocessing steps, including the strategic use of TF-IDF, are meticulously designed to optimize our dataset for the retrieval model. This approach is a cornerstone of our effort to achieve a grade **B+ level** of accomplishment, ensuring that our information retrieval system is not only accurate but also contextually relevant to user queries.

3. Results Retrieval

To effectively retrieve information that aligns closely with user queries, our system employs a cosine similarity-based approach, a key component of our retrieval methodology. Cosine similarity is a metric used to measure how similar two documents are irrespective of their size. Mathematically, it calculates the cosine of the angle between two vectors projected in a multi-dimensional space. In the context of our system, these vectors represent the user query and the documents in our dataset, respectively. By using cosine similarity, we can quantify the similarity between a query and each document, providing a robust method to determine relevance.

The retrieval process involves several key steps:
Query Vectorization: Each user query is converted into a vector using the same TF-IDF approach as our dataset. This ensures that the query and the documents are in a comparable format for analysis.
Document Sorting: Documents are then ranked based on their cosine similarity scores with the vectorized query. This ranking effectively identifies which documents are most relevant to the user's search.
Top Results Selection: Finally, the system selects and presents the top 5 documents with the highest relevance scores. This selection is based on the calculated similarity, ensuring that the most pertinent and useful information is provided to the user.

This approach, incorporating the nuanced application of cosine similarity, aligns with our goal of achieving a grade **A- level** of accomplishment.

4. Result Summarization

The final stage in our information retrieval process involves summarizing the retrieved results:
Document Aggregation: Combining full texts of the top 5 documents ensures comprehensive coverage of relevant information.
Sentence Ranking and Selection: In the initial stages, we incorporated Lexrank for summarization , LexRank is an unsupervised approach that uses graph-based centrality scoring to identify the most significant sentences in a text, akin to how Google's PageRank algorithm ranks web pages. By applying LexRank, we could effectively highlight sentences that are most relevant and informative in relation to the user query.

But, In our updated text summarization approach, we utilize the T5 (Text-to-Text Transfer Transformer) model, which has been pre-trained on a multi-task mixed dataset and is particularly suited for generating summaries. This model outperforms the previous LexRank approach by utilizing a more advanced neural network architecture that has been fine-tuned on a diverse range of language understanding tasks, including summarization.

Concise Presentation: The summarized result, a blend of information from the top documents, is presented in a user-friendly, concise format and the top 5 documents ranked on its relevance is also provided as shown below in the results generated by our model.

This summarization process is vital for providing users with clear, relevant information efficiently, reflecting the innovative nature of our project and echoes our objective of achieving a grade **A+ level** accomplishment.

Query Analysis

Our project is tailored for a closed-domain system, focusing on rare diseases. Users can input queries about symptoms, medications, recent research, affected organs, and diagnoses or treatments related to rare diseases. The system is primarily designed for researchers and medical professionals seeking detailed information on rare diseases. Each query undergoes vectorization using TF-IDF to align with our dataset format. We then utilize cosine similarity to measure the relevance of the dataset documents to the query, ensuring accurate and pertinent information retrieval.

Results Analysis

In our project, the relevance of results is strictly aligned with the specificity and accuracy of rare diseases. The results are considered relevant if they precisely address the user's query, which may encompass symptoms, treatments, recent research, and organ impact of specific rare diseases.

Our system is designed to retrieve a set volume of results per query to maintain a balance between comprehensiveness and user manageability. For each query, the top 5 documents from our dataset, ranked by their relevance scores, are presented. This number was chosen to provide users with a focused yet diverse range of information, enabling them to gain a comprehensive understanding of the topic without being overwhelmed by excessive data. The retrieved documents are not only ranked by relevance but also summarized.

Evaluation Metrics

Accuracy is the key metric used to evaluate the performance of our information retrieval system. It is specifically defined and measured based on how well the system's results align with a predefined set of highly relevant documents, referred to as the ground truth.

For each example query, we have identified the top 3 documents that represent the most relevant and accurate information regarding the query. These top 3 documents constitute our ground truth against which the system's accuracy is measured. The accuracy of our model is determined by

comparing its results with the ground truth. We anticipate the model to predict at least 2 out of 3 documents from the ground truth. This criterion reflects the model's ability to successfully identify and prioritize the most relevant documents from the dataset in response to a user query.

Performance Analysis

The model's performance, with an accuracy rate of around 72%, demonstrates its strong capability in accurately identifying relevant documents from our dataset. This rate indicates that in approximately 72% of instances, the model successfully predicts at least 2 out of the top 3 relevant documents within its top 5 results. While this performance is commendable, especially in the complex field of rare disease information retrieval, there is still room for improvement. The high accuracy rate positively impacts user experience, providing reliable and pertinent information, thus increasing the system's trustworthiness and usability. Future enhancements, such as refining semantic analysis and algorithm adjustments, are anticipated to further improve this accuracy. Overall, the model’s effectiveness in rare disease information retrieval not only aids in better understanding these conditions but also enhances research and treatment approaches in this specialized medical field.

Conclusion

In conclusion, our project successfully developed a sophisticated model designed to navigate the complex landscape of rare disease information retrieval. By employing techniques such as TF-IDF vectorization, cosine similarity-based document ranking, and T5 transformer model for summarization, we have achieved a good degree of accuracy and relevance in our information retrieval system. The accuracy rate and the relevance scores for retrieved documents underscores the system's effectiveness in aligning closely with user queries and providing pertinent information. This project not only marks a significant stride in the realm of medical information retrieval but also sets a precedent for future innovations in the field. Going forward, we anticipate continuous improvements and refinements to further enhance the system's capabilities and user experience, solidifying our commitment to bridging the information gap in rare disease research and treatment.

Challenges and Future Works

Annotations by medical experts in the field of rare diseases will improve the annotation of our ground truth documents. This will ensure that the annotated data is accurate, comprehensive, and clinically relevant, enhancing the quality of our dataset. To overcome the limitations posed by our current computational resources, we will explore more sophisticated text embedding techniques. These techniques, such as deep learning models or advanced natural language processing algorithms, can provide more nuanced and contextually aware representations of our dataset. We aim to optimize our computational resource usage to handle our large dataset more efficiently. This may involve leveraging cloud computing services or optimizing our existing algorithms for better performance. We will continue to refine and expand our dataset, ensuring that it remains up-to-date with the latest research and developments in the field of rare diseases. This ongoing process will help in maintaining the relevance and accuracy of our system.

By focusing on these areas, we aim to not only overcome the current challenges but also significantly advance the capabilities and usefulness of our information retrieval system in the medical field.

Contribution breakdown

Manasa: Data annotation, result summarization and presentation, query generation, performance analysis, documentation.
Chintan: Data acquisition, creation and preprocessing, query generation and results retrieval/ranking relevant documents, documentation.

References

1. **Link to GitHub Repo:** https://github.com/chintansa/CS6200-IR-Project
2. **Link to Dataset created:** IR-CS-6200-Final_project_files
3. **Orphanet**: https://www.orpha.net/consor/cgi-bin/index.php
4. **PubMed**: https://pubmed.ncbi.nlm.nih.gov/
5. **Tf-idf:** https://en.wikipedia.org/wiki/Tf%E2%80%93idf
6. **Cosine similarity:** https://en.wikipedia.org/wiki/Cosine_similarity
7. **LexRank:** https://en.wikipedia.org/wiki/Automatic_summarization
8. **T5 transformer model**: AutoModels — transformers 3.0.2 documentation

Queries and Results

The ground truth indicates the PMIDs(ID of the article) here

Query: Gene editing prospects in treating rare genetic diseases
Narrative: A geneticist is investigating the potential of gene-editing technologies as treatment options for patients with rare genetic disorders.
Summary:
'gene editing is a powerful tool for genome and cell engineering. exemplified by CRISPR-Cas, gene editing could cause DNA damage. catalytically inactive dCas9 promotes knock-in of long sequences in mammalian cells. dCas9-SSAP editor has low on-target errors and minimal off-target effects. dCas9-SSAP is effective for inserting kilobase-scale sequences '

Relevant Documents:

	PMID	title	article_url	relevance_score
0	35145221	dCas9-based gene editing for cleavage-free gen...	https://doi.org/10.1038/S41556-021-00836-1	0.207978
1	30524313	New Approaches to Tay-Sachs Disease Therapy.	https://doi.org/10.1038/S41586-019-1711-4	0.194631
2	31634902	Search-and-replace genome editing without doub...	https://doi.org/10.1038/S41586-019-1711-4	0.194631
3	35741383	Gene Editing-Based Technologies for <i>Beta-he...	https://doi.org/10.1056/Nejmoa1705342	0.181994
4	37298481	Precision Editing as a Therapeutic Approach fo...	https://doi.org/10.1080/20009666.2018.1536241	0.173292

Ground Truth: 35145221, 37298481, 35098209

Query: Robotic surgery in the treatment of rare urogenital diseases
Narrative: A urological surgeon is researching the application of robotic-assisted surgical techniques in the treatment of rare urogenital diseases, aiming to understand advanced procedures.
Summary:
'robotic surgery has shown clear utility and advantages in the adult population. but its role in pediatrics remains controversial. robotic pyeloplasty is a standard of care in older children and has even been performed in infants and re-do surgery. future advances in robotics will help to advance the field of robotic surgery in pediatric urology. back to mail online home. back to the page you came from. back to the page you came from.'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	29264208	Robotic surgery in pediatric urology.	https://doi.org/10.1016/J.Ajur.2016.06.002	0.403642
1	35038623	Intensified Hyposensitization Is an Effective ...	https://doi.org/10.1016/J.Fertnstert.2019.11.021	0.187432
2	35159098	A Comprehensive Commentary on the Multilocular...	https://doi.org/10.3348/Jksr.2021.0022	0.175941
3	27833904	Surgical Treatment after Failed Primary Correc...	https://doi.org/10.1002/Uog.7721	0.167779
4	35440058	Molecular cytogenetic characterization of part...	https://doi.org/10.1159/000511972	0.166603

Ground Truth: 27833904, 35038623, 29610640

Query: Role of immunotherapy in treating rare endocrine disorders
Narrative: An endocrinologist is exploring the effectiveness of immunotherapy in managing rare endocrine disorders and mechanism of action of immunotherapeutic agents in these conditions.
Summary:
'the therapeutic landscape for non-melanoma skin cancer has expanded with the development of effective and targeted immunotherapy. in this review, we discuss relevant ophthalmic findings associated with key disorders of the pancreas, thyroid gland, and hypothalamic-pituitary axis. diabetes mellitus (DM) is the leading cause of blindness among adults under 75 years of age. ophthalmic retinopathy (DR) has significant predictive value for cardiovascular disease and mortality in patients with'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	34448958	Immunotherapy for Non-melanoma Skin Cancer.	https://doi.org/10.2340/00015555-0359	0.345397
1	29184811	Dermatologic manifestations of endocrine disor...	https://doi.org/10.21037/Tp.2017.09.08	0.271060
2	31687637	The Natural History of a Man With Ovotesticula...	https://doi.org/10.1210/Js.2019-00241	0.250915
3	36613725	Childhood Hypophosphatasia Associated with a N...	https://doi.org/10.1210/Js.2017-00307	0.250636
4	34884378	Hypophosphatasia.	https://doi.org/10.1210/Js.2017-00307	0.250636

Ground Truth: 34448958, 29184811, 34884378

Query: Prenatal diagnostic methods for rare teratologic disorders
Narrative: A reseacher seeks information on the latest advancements in prenatal diagnostic techniques that can detect rare congenital (teratologic) disorders.
Summary:
'angelman syndrome (AS) is a rare neurodevelopmental disorder caused by mutation or deletion of the maternally inherited UBE3A allele. genetic tests can detect the chromosome 15q11-q13 deletion that is the most common cause of AS. prenatal testing combined with prenatal treatment has the potential to revolutionize how clinicians detect and treat babies before they are symptomatic. this pioneering prenatal treatment path for AS will lay the foundation for treating other syndromic neurodevelopmental disorders'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	27590389	Prenatal diagnosis of partial monosomy 5p (5p1...	https://doi.org/10.1016/J.Tjog.2016.06.014	0.156389
1	31490639	Prenatal treatment path for angelman syndrome ...	https://doi.org/10.1002/Aur.2203	0.151882
2	27208505	A diagnostic approach to mild bleeding disorders.	https://doi.org/10.1111/Jth.13368	0.140649
3	22030049	Prenatal diagnosis of mosaic trisomy 8: clinic...	https://doi.org/10.1016/J.Tjog.2011.07.013	0.130093
4	36034547	Prenatal ultrasound phenotypic and genetic eti...	https://doi.org/10.1016/J.Ajhg.2010.06.010	0.128718

Ground Truth: 27126916, 33745446, 28184291

Query: Renal replacement therapy options for rare renal diseases
Narrative: A nephrologist is looking for information on renal replacement therapies, specifically for patients suffering from rare kidney diseases, aiming to improve their treatment strategies.
Summary:
'atherosclerotic renal artery stenosis is the leading cause of secondary hypertension. it may lead to resistant (refractory) hypertension, progressive decline in renal function, and cardiac destabilization syndromes. the best strategy to approach the treatment of atherosclerotic renal artery stenosis is to revascularize patients with renal artery stenosis. the authors review techniques to optimize patient selection, to minimize procedural complications, and to facilitate durable pat'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	21071977	Tuberous sclerosis complex renal disease.	https://doi.org/10.1159/000320891	0.422263
1	30898248	When and How Should We Revascularize Patients ...	https://doi.org/10.1016/J.Jcin.2018.10.023	0.363647
2	33603889	Management of acute intradialytic cardiovascul...	https://doi.org/10.1016/J.Jcin.2018.10.023	0.362259
3	30274631	Renal Manifestations of Inflammatory Bowel Dis...	https://doi.org/10.1016/J.Rdc.2018.06.007	0.358660
4	25316474	Renal artery stenosis in association with cong...	https://doi.org/10.1038/S41375-021-01290-6	0.358251

Ground Truth: 21071977, 308998248, 29488740

Query: Biologic drug therapies for rare skin diseases
Narrative: A dermatologist is researching the use of biologic drugs in treating rare skin diseases, wanting to understand the latest biologic treatments available, and prognosis for patients receiving these therapies.
Summary:
'skin cancer has been the leading type of cancer worldwide. skin fibrosis is characterized by excessive fibroblast proliferation and extracellular matrix deposition in the dermis. effective anti-scarring therapeutics remain an unmet need, says dr. sanjay gupta. gupta: the mortality rate of severe drug eruptions can reach up to 50% if not treated early. gupta: a multidisciplinary approach is required for acute management '

Relevant Documents:

	PMID	title	article_url	relevance_score
0	35127174	Skin cancer biology and barriers to treatment:...	https://doi.org/10.1016/J.Jare.2021.06.014	0.223879
1	20065636	Canakinumab.	https://doi.org/10.1038/Nm.3804	0.212849
2	34273058	Current Perspectives on Severe Drug Eruption.	https://doi.org/10.1007/S12016-021-08859-0	0.211538
3	31493000	The IL-4/IL-13 axis in skin fibrosis and scarr...	https://doi.org/10.1007/S00403-019-01972-3	0.203446
4	34641447	Controlled Drug Delivery Systems: Current Stat...	https://doi.org/10.1016/J.Chest.2021.10.010	0.202744

Ground Truth: 31493000, 34641447, 17187155

Query: Symptoms of rare respiratory diseases
Narrative: The user needs to find a comprehensive list of symptoms associated with rare respiratory diseases.
Summary:
‘costello syndrome (CS) is a multisystem disorder caused by heterozygous germline mutations in the HRAS proto-oncogene. in the neonatal period, respiratory complications are seen in approximately 78% of patients with transient respiratory distress reported in 45% of neonates. other more specific respiratory diagnoses were reported in 62% of patients, the majority of which comprised disorders of the upper and lower respiratory tract. this review summarizes neonatal respiratory therapy's advances and available strategies.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	28877977	Rare pulmonary diseases: a common fight.	https://doi.org/10.1038/S41576-022-00478-5	0.466163
1	27102959	Respiratory system involvement in Costello syn...	https://doi.org/10.1002/Ajmg.A.37655	0.410557
2	35382987	New developments in neonatal respiratory manag...	https://doi.org/10.1016/J.Pedneo.2022.02.002	0.285090
3	32016537	Respiratory drive in the acute respiratory dis...	https://doi.org/10.1007/S00134-020-05942-6	0.273988

Ground Truth: 35382987, 34342628, 36588756

Query: Fertility treatment outcomes in rare infertility cases
Narrative: The user is compiling success rates of various treatments in patients with rare causes of infertility.
Summary:
'Approximately half of male factor infertility cases have no known cause. but, it is likely that the majority of idiopathic male factor infertility cases have some unidentified genetic basis. Approximately 15% of couples worldwide are affected with infertility, attributed to a male co-factor in about half of the cases. most studies evaluate a single gene, an approach that is very inefficient in the context of male infertility.'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	24574159	Genetic susceptibility to male infertility: ne...	https://doi.org/10.1111/J.2047-2927.2014.00188.X	0.395717
1	32622407	Fertility issues and pregnancy outcomes in Tur...	https://doi.org/10.1016/J.Fertnstert.2020.03.002	0.377151
2	31355535	Clinical and molecular characterization of Y m...	https://doi.org/10.1111/Andr.12686	0.339412
3	33071633	Genetic disorders and male infertility.	https://doi.org/10.1002/Rmb2.12336	0.336226

Ground Truth: 24574159, 36668910, 35215786

Query: Long-term outcomes of rare systemic diseases in childhood
Narrative: The user is gathering data on the long-term health outcomes for children diagnosed with rare systemic diseases.
Summary:
'childhood glaucoma is classified in primary and secondary congenital glaucoma. CYP1B1 gene mutations seem to account for 87% of familial cases. childhood absence epilepsy (CAE) is considered easily manageable with medication. childhood absence epilepsy (epi) is most frequent in school-aged girls and is most frequent in school-aged girls. epi is a rare chromosomal disorder with distinctive phenotypic expressivity.'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	33466173	Prenatal diagnosis and ultrasonographic findin...	https://doi.org/10.1016/J.Amjms.2021.09.004	0.202516
1	26451378	Rare Diseases Leading to Childhood Glaucoma: E...	https://doi.org/10.2478/Bjmg-2019-0024	0.181602
2	31733607	B cell targeted therapies in autoimmune disease.	https://doi.org/10.1016/J.Coi.2019.09.004	0.177663
3	28325560	Long-term prognosis of childhood absence epile...	https://doi.org/10.1016/J.Nrl.2016.12.005	0.170796
4	38045990	Adult Phenotype of SYNGAP1-DEE.	https://doi.org/10.1016/J.Seizure.2009.04.004	0.170199

Ground Truth: 38045990, 31733607, 26451378

Query: Diagnostic criteria for rare hepatic diseases
Narrative: A hepatologist is looking for established diagnostic criteria for rare liver diseases to assist with a challenging patient case.
Summary:
'hepatic processes are dysregulated in type 1 and type 2 diabetes mellitus. this imbalance contributes to hyperglycaemia in the fasted and postprandial states. in this review, we discuss the in vivo regulation of these hepatic glucose fluxes. we also highlight the importance of indirect (extrahepatic) control of hepatic gluconeogenesis and direct (hepatic) control of hepatic glycogen metabolism.'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	28731034	Regulation of hepatic glucose metabolism in he...	https://doi.org/10.1038/Nrendo.2017.80	0.321340
1	31148902	Autoimmune hepatitis and IgG4-related disease.	https://doi.org/10.3748/Wjg.V25.I19.2308	0.307815
2	36031652	Infantile hepatic hemangioma and hepatic mesen...	https://doi.org/10.1002/Ajmg.A.62767	0.281110
3	34928431	Challenges in the diagnosis of neurofibromatos...	https://doi.org/10.1007/S00439-021-02410-Z	0.279409

Ground Truth: 28731034, 31148902, 36031652

Query
What are the causes of IDA?
Narrative
The query seeks comprehensive information on the factors leading to Iron Deficiency Anemia (IDA).
Summary
‘iron deficiency anemia (IDA) is the most prevalent and treatable form of anemia worldwide. treatment landscape for many causes of IDA is shifting toward more abundant use of intravenous (IV) iron. anemia is a global health problem affecting one-third of the world population, and half of the cases are due to iron deficiency (ID) a rare genetic form called iron-refractory iron deficiency anemia exists.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	31715055	The role of iron repletion in adult iron defic...	https://doi.org/10.1111/Ejh.13345	0.566807
1	34007416	Acquired Refractory Iron Deficiency Anemia.	https://doi.org/10.1016/J.Clml.2018.11.019	0.336498

Ground Truth: 31715055, 34007416, 27126916

Query
Diagnosis of DILI
Narrative
The query focuses on understanding the methods and criteria used to diagnose Drug-Induced Liver Injury (DILI), including specific tests, markers, and clinical assessment procedures.
Summary
‘drug-induced liver injury (DILI) is an adverse reaction to drugs or other xenobiotics. it is a relative rare hepatic disorder but can be severe and, in some cases, fatal. the diagnosis of DILI relies on the exclusion of other aetiologies of liver disease. a number of clinical variables, validated in prospective cohorts, can be used to predict a more severe outcome.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	31439850	Drug-induced liver injury.	https://doi.org/10.1038/S41572-019-0105-0	0.676320
1	35979160	Drug-induced autoimmune hepatitis: A minireview.	https://doi.org/10.3748/Wjg.V28.I24.2654	0.347729
2	34948375	Pathogenesis of Autoimmune Hepatitis-Cellular ...	https://doi.org/10.3350/Cmh.2020.0189	0.083325

Ground Truth: 31439850, 20113479, 33718381

Query
LYG disorder
Narrative
The query aims to obtain a concise yet comprehensive understanding of Lymphomatoid Granulomatosis (LYG), focusing on its definition, symptoms, causes, and any known treatment options.
Summary
'lymphomatoid granulomatosis (LYG) is a rare Epstein-Barr virus (EBV) driven B-cell lymphoproliferative disease (LPD) clinically, LYG universally appears to be a rare form of lymphoid granulomatosis. lyg is graded by the number and density of EBV+ atypical B cells. clinically, LYG universally appears to be a rare form of lympho'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	32107539	Pathobiology and treatment of lymphomatoid gra...	https://doi.org/10.1182/Blood.2019000933	0.114715
1	12358356	Primary pulmonary lymphoma.	https://doi.org/10.1182/Blood.2019000933	0.114715

Ground Truth: 32107539, 12358356, 36389673

Query
Are women affected by Turner Syndrome?
Narrative
The query seeks to confirm whether Turner Syndrome exclusively affects women, exploring its genetic basis and how it manifests specifically in the female population.
Summary
'clinical manifestations of Turner syndrome include abnormalities of the skeletal, cardiovascular and lymphatic systems. ovarian function sufficient to result in puberty is uncommon in women with Turner syndrome. there are several options for women who desire to expand their families. despite improved expectations for fertility in individuals with Turner syndrome, pregnancy-related mortality remains higher than in the general population. despite improved expectations for fertility in individuals with Turner syndrome, pregnancy-related mortality remains high.'

Relevant Documents:

	PMID	title	article_url	relevance_score
0	26568488	Reproductive Issues in Women with Turner Syndr...	https://doi.org/10.1016/J.Ecl.2015.07.004	0.453546
1	28228961	Concurrent Van der Woude syndrome and Turner s...	https://doi.org/10.3238/Arztebl.2020.0751	0.422633
2	32622407	Fertility issues and pregnancy outcomes in Tur...	https://doi.org/10.1016/J.Fertnstert.2020.03.002	0.376837
3	17708142	Turner syndrome: diagnosis and management.	https://doi.org/10.1053/J.Ajkd.2020.08.012	0.367579

Ground Truth: 26568488, 4086312, 31546693

Query

Clinical case on neuromuscular scoliosis

Narrative

The query seeks insights into a specific patient case involving neuromuscular scoliosis, focusing on aspects like diagnosis, treatment approaches, patient management, and outcomes.

Summary

‘a 12-year-old girl with shprintzen-goldberg syndrome was hospitalized for a surgical treatment. she was diagnosed with 80° symptomatic scoliosis in T10-L5. lower incidences of intraspinal abnormalities in patients with early onset idiopathic scoliosis have been observed in the uk. fusionless surgery can be performed to delay final fusion until the patient is older.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	33299628	Scoliosis in Shprintzen-Goldberg Syndrome.	https://doi.org/10.1038/S41436-019-0609-8	0.305843
1	28566405	Intraspinal anomalies in early-onset idiopathi...	https://doi.org/10.1302/0301-620X.99B6.Bjj-201...	0.275230
2	25623270	Early-onset scoliosis: current treatment.	https://doi.org/10.1016/J.Otsr.2014.06.032	0.263347
3	28356803	Superior mesenteric artery syndrome.	https://doi.org/10.6001/Actamedica.V23I3.3379	0.258420

Ground Truth: 25623270, 28356803, 33299628

Query

Life-threatening cardiovascular or cerebrovascular complications

Narrative

This query seeks information on severe cardiovascular or cerebrovascular complications that pose a life-threatening risk, focusing on their causes, symptoms, diagnosis, and treatment options.

Summary

‘the displacement of a third molar is a rare occurrence, but it could lead to serious and/or life threatening complication. the bosley-salih-alorainy syndrome (BSAS) variant of the congenital human HOXA1 syndrome results from autosomal recessive truncating HOXA1 mutations. we describe the currently recognized spectrum of ocular motility, inner ear malformations, cerebrovascular anomalies.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	35678401	Vascular Cognitive Impairment and Dementia.	https://doi.org/10.1212/Con.0000000000001124	0.257067
1	36300346	Association of Vascular Risk Factors and Genet...	https://doi.org/10.1001/Jamaneurol.2022.3832	0.218298
2	30697395	Immediate or delayed retrieval of the displace...	https://doi.org/10.4317/Jced.55379	0.175544
3	17875913	Clinical characterization of the HOXA1 syndrom...	https://doi.org/10.1002/Cncr.29920	0.169273

Ground Truth: 35678401, 30697395, 36300346

Query

What are some rare respiratory disorders?

Narrative

The query "seeks to identify and provide brief descriptions of various uncommon or less widely-known respiratory diseases, focusing on their names and basic characteristics.

Summary

‘costello syndrome (CS) is a multisystem disorder caused by heterozygous germline mutations in the HRAS proto-oncogene. in the neonatal period, respiratory complications are seen in approximately 78% of patients with transient respiratory distress reported in 45% of neonates. in cases of prenatally diagnosed CS, the high incidence of respiratory complications in the neonatal period should prompt anticipatory guidance and development of a postnatal management plan.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	27102959	Respiratory system involvement in Costello syn...	https://doi.org/10.1002/Ajmg.A.37655	0.451983
1	28877977	Rare pulmonary diseases: a common fight.	https://doi.org/10.1038/S41576-022-00478-5	0.402232
2	35382987	New developments in neonatal respiratory manag...	https://doi.org/10.1016/J.Pedneo.2022.02.002	0.332260
3	32016537	Respiratory drive in the acute respiratory dis...	https://doi.org/10.1007/S00134-020-05942-6	0.319322

Ground Truth: 27102959, 28877977, 3040392

Query

Types of infertility

Narrative

The query aims to gather detailed information about the different classifications of infertility, focusing on their causes, characteristics, and how they affect both males and females.

Summary

/Approximately 15% of couples worldwide are affected with infertility, attributed to a male co-factor. azoospermia is the most challenging topic associated with infertility treatment. genetic abnormalities are found in 10%-20% of patients showing severe spermatogenesis disorders. azoospermia is a rare form of infertility that can be treated with a combination of drugs and hormone therapy.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	24574159	Genetic susceptibility to male infertility: ne...	https://doi.org/10.1111/J.2047-2927.2014.00188.X	0.558361
1	31355535	Clinical and molecular characterization of Y m...	https://doi.org/10.1111/Andr.12686	0.473474
2	33071633	Genetic disorders and male infertility.	https://doi.org/10.1002/Rmb2.12336	0.454825
3	23503951	The epidemiology and etiology of azoospermia.	https://doi.org/10.1038/Npjgenmed.2016.23	0.431037

Ground Truth: 31355535, 24574159,

Query
Rare renal diseases
Narrative
The query aims to gather information on uncommon or less frequently occurring kidney diseases, focusing on their types, symptoms, causes, diagnosis, and treatment options.
Summary
‘patients with tuberous sclerosis complex frequently have significant renal involvement. 4% to 23% of inflammatory bowel disease (ibd) patients have renal and urinary involvement. there are no established guidelines for the optimal screening and monitoring of renal function in IBD patients. renal artery stenosis is the leading cause of secondary hypertension and may lead to resistant hypertension (HTN), progressive decline in renal function, and cardiac destabilization syndromes.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	21071977	Tuberous sclerosis complex renal disease.	https://doi.org/10.1159/000320891	0.447411
1	30274631	Renal Manifestations of Inflammatory Bowel Dis...	https://doi.org/10.1016/J.Rdc.2018.06.007	0.376579
2	25316474	Renal artery stenosis in association with cong...	https://doi.org/10.1038/S41375-021-01290-6	0.364024
3	30898248	When and How Should We Revascularize Patients ...	https://doi.org/10.1016/J.Jcin.2018.10.023	0.362665
4	33603889	Management of acute intradialytic cardiovascul...	https://doi.org/10.1016/J.Jcin.2018.10.023	0.361280

Ground Truth: 21071977, 30274631, 25316474

Query
Seasonal changes and allergy
Narrative
The query aims to explore how seasonal variations impact the occurrence and severity of allergic reactions, focusing on identifying specific allergens prevalent in different seasons and their effects on susceptible individuals.
Summary
‘a comprehensive review of the literature underlining all the forms classified as ocular allergy. ocular allergy represents one of the most common conditions encountered by allergists and ophthalmologists. pharmacological agents have improved the efficacy and safety of ocular allergy treatment. a single cause of this increase cannot be pinpointed and experts are considering factors including genetics, air pollution in urban areas, pets, and early childhood exposure.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	23497516	Allergic conjunctivitis: a comprehensive revie...	https://doi.org/10.1186/1824-7288-39-18	0.229495
1	26563668	International Consensus Document (ICON): Commo...	https://doi.org/10.1016/J.Jaip.2015.07.025	0.226066
2	28748365	Urticaria and Angioedema: an Update on Classif...	https://doi.org/10.1007/S12016-017-8628-1	0.139066

Ground Truth: 23497156, 31207234, 34943306

Query

Inflammatory arthralgia case

Narrative

The query seeks specific information regarding the characteristics, causes, and possible treatments of inflammatory arthralgia, focusing on symptoms, diagnostic methods, and management strategies for this joint condition.

Summary

‘the NLRP3 inflammasome is a multimeric cytosolic protein complex. it assembles in response to cellular perturbations, leading to the activation of caspase-1. inflammatory cytokines contribute to the development of systemic low-grade inflammation. NLRP3 inflammasome-mediated inflammatory cytokines play dual roles in mediating human disease. it has a beneficial role in numerous infectious diseases and some cancers.’

Relevant Documents:

	PMID	title	article_url	relevance_score
0	33707781	NLRP3 inflammasome in cancer and metabolic dis...	https://doi.org/10.1038/S41590-021-00886-5	0.171689
1	32818434	COVID-19 and multisystem inflammatory syndrome...	https://doi.org/10.1016/S1473-3099(20)30651-4	0.152896
2	33640145	Multisystem inflammatory syndrome in children ...	https://doi.org/10.1016/J.Vaccine.2021.01.054	0.142644

Ground Truth: 33640145, 3281834, 33707781