Accuracy of generative artificial intelligence models in differential diagnoses of familial Mediterranean fever and deficiency of Interleukin-1 receptor antagonist

Joshua Pillai [a,*], Kathryn Pillai [b]

[a] *Department of Biological Sciences, Irvine Unified School District. 5050 Barranca Parkway, Irvine, 92604, CA, USA*
[b] *Department of Medical Education, California University of Science and Medicine, School of Medicine. 1501 Violet St, Colton, CA, USA*

ARTICLE INFO

ABSTRACT

With the increasing development of artificial intelligence, large language models (LLMs) have been utilized to solve problems in natural language processing tasks. More recently, LLMs have shown unique potential in numerous applications within medicine but have been particularly investigated for their ability in clinical reasoning. Although the diagnostic accuracy of LLMs in forming differential diagnoses has been reviewed in general internal medicine applications, much is unknown in autoinflammatory disorders. From the nature of autoinflammatory diseases, forming a differential diagnosis is challenging due to the overlapping symptoms between disorders and even more difficult without genetic screening. In this work, the diagnostic accuracy of the Generative Pre-Trained Transformer Model-4 (GPT-4), GPT-3.5, and Large Language Model Meta AI (LLaMa) were evaluated in clinical vignettes of Deficiency of Interleukin-1 Receptor Antagonist (DIRA) and Familial Mediterranean Fever (FMF). We then compared these models to a control group including one internal medicine physician. It was found that GPT-4 did not significantly differ in correctly identifying DIRA and FMF patients compared to the internist. However, the physician maintained a significantly higher accuracy than GPT-3.5 and LLaMa 2 for either disease. Overall, we explore and discuss the unique potential of LLMs in diagnostics for autoimmune diseases.

1. Introduction

The rapid development of artificial intelligence (AI) has led to the creation of large language models (LLMs). The foundations of LLMs are based on deep learning architectures trained on large datasets involving articles, websites, and other forms of text. This has led to the formation of commercial models such as OpenAI's Generative Pre-Trained Transformer-3 (GPT-3) and Meta's Large Language Model Meta AI (LLaMa) to solve natural language processing tasks [1,2]. Within medicine, these models have been thoroughly investigated in numerous applications. For instance, they have helped automating electronic health records and assisting in systematic reviews [3,4]. More recently, though, LLMs have been thoroughly investigated for complex diagnostic reasoning.

Given the recent success of LLMs in scoring near or above the passing threshold of the United States Medical Licensing Exam [5], current studies have evaluated the accuracy of these models in forming a differential diagnosis [6–8]. In a previously reported study by Kottlors et al.

[6], GPT-4 scored acceptable to clinical professionals in imaging patterns (68.8 %–93.8 %, respectively). In another investigation by Kanjee et al. [7], authors found that GPT-4 was able to identify a correct diagnosis in 64 % of cases and 39 % as its primary diagnosis.

However, even with the thorough testing of LLMs in determining their diagnostic accuracy, much is still unknown of their abilities in challenging cases. In particular, autoinflammatory disorders are complex cases for their nature of overlapping symptoms such as systemic inflammation, increased acute phase reactants, acute arthritis, and recurrent fevers [9]. Additionally, without thorough genetic screening, forming a diagnosis of these diseases is even more difficult. In this work, we assess the accuracy of GPT-4, compared with GPT-3.5 and LLaMa 2, in forming a diagnosis for the Deficiency of Interleukin-1 Receptor Antagonist (DIRA) and Familial Mediterranean Fever (FMF).

DIRA is an ultra-rare autoinflammatory disorder that is caused by an autosomal recessive mutation in the *IL1RN* gene and therefore cannot encode IL-1Ra proteins [9]. This leads to abnormal activity of IL-1β and IL-1α and causes an inflammatory state [9]. The primary symptoms of

List of abbreviations:

AI	Artificial Intelligence
DIRA	Deficiency of Interleukin-1 Receptor Antagonist
FMF	Familial Mediterranean Fever
GPT-4	Generative Pre-Trained Transformer-4
GPT-3	Generative Pre-Trained Transformer-3
LLaMa	Large Language Model Meta AI
LLM	Large Language Model

DIRA include systemic inflammation, multifocal osteomyelitis, and elevated acute phase reactants [10].

In comparison, FMF is an autoinflammatory disorder caused by an autosomal recessive mutation in the *MEFV* gene, resulting in lower production of pyrin [11]. This eventually causes an increase in IL-1β and forms an inflammatory response. The common symptoms of FMF are recurrent episodes of fever, seriotitis, abdominal pain, chest pain, and skin lesions [11].

Within this study, we prepared clinical vignettes of DIRA and FMF patients obtained from the PubMed search engine. Then, we evaluated and recorded the results from GPT-4, GPT-3.5, LLaMa 2, and compared them to an American Board Certified Internal Medicine Physician. Subsequently, we analyzed and discussed the scope of the results.

2. Materials and methods

2.1. Study design

Diagnostic accuracies were assessed of GPT-3.5 (OpenAI, San Francisco, CA, USA), GPT-4 (OpenAI, San Francisco, CA, USA), and LLaMa 2 (Meta AI, Menlo Park, CA, USA) models. This study was conducted at the California University of Science and Medicine, School of Medicine, Department of Medical Education, Colton, California, USA. A blind test was conducted on a certified internist to evaluate the accuracy compared to AI models. The clinical vignettes of DIRA and FMF were obtained from published manuscripts, and therefore did not require individual consent or approval from an Institutional Review Board.

GPT-3.5 (chat.openai.com) was accessed on August 11, 2023, GPT-4 (chat.openai.com) on August 12, 2023, and LLaMa 2 (labs.perplexity.ai, 70 b model) on August 13, 2023. For all models, there were no hyperparameters specifically tuned for clinical diagnosis.

2.2. Clinical vignettes

All case vignettes were obtained from the PubMed database. We used the following key-word index: "Familial Mediterranean Fever" or "FMF" and "Deficiency of Interleukin-1 Receptor Antagonist" or "DIRA." Language restrictions were applied to only English. Gray literature was not included in the search, and all article types included were case reports. We obtained a total of twenty written vignettes for both DIRA [12–23] and FMF [24–42], respectively (*n* = 40).

For all written clinical vignettes, the format followed the history of present illness including all of the following information: Age, Sex, Clinical Presentation, Related Family and Patient History, Physical Examination, and any Laboratory Results. Each vignette had only one correct diagnosis. However, to determine the accuracy of models, we allowed up to 10 differential diagnoses and sorted them into 3 categories: primary choice (top choice), top five, and top ten. All 40 patient vignettes have been provided in Supplementary Material 1. We then analyzed each element, including age, sex, chief complaint, consanguinity, and laboratory results.

2.3. Data extraction

Due to the potential of LLMs to recall previous information within the same conversation, each vignette was shown in differing conversations. Additionally, the order of vignettes displayed were randomized through the covariate adaptive randomization technique through GraphPad Prism (version 10.0.2) [43].

Before extraction, J.P·P typed the following text into all models: "I am conducting an experiment on the accuracy of differential diagnoses in clinical cases to assess how you compare with other models. I am going to list a clinical case below, and want you to give me your top ten diagnoses. Do not list any explanations. Please only list ten diagnoses. Here is the case below: (clinical vignette)." The models then listed their ten diagnoses and the data was recorded. For the control group, the internist was given the same randomized vignettes tested on the AI models. The reference section was omitted for both groups.

2.4. Statistical analyses

Data was collected using Excel and statistical analysis was completed on GraphPad Prism (version 10.0.2). Due to the moderate sample size of each group, a Fisher's Exact test was conducted on categorical data. For continuous datasets, a Mann Whitney *U* Test was conducted to evaluate significance. Unless stated otherwise, a *P*-value below 0.05 was considered statistically significant.

3. Results

Within this pilot study, we report 40 clinical vignettes tested on GPT-4, GPT-3.5, and LLaMa 2 compared to an internal medicine physician. The results of each model's diagnostic accuracy has been displayed in Fig. 1.

Overall, all models had a higher accuracy identifying FMF compared to DIRA. The internist identified 90 %, GPT-4 identified 65 %, GPT-3.5 identified 60 %, and LLaMa 2 identified 17.6 % of FMF patients. Conversely, the internist identified 60 %, GPT-4 identified 30 %, GPT-3.5 identified 25 %, and LLaMa 2 identified 0 % of DIRA patients.

The internist performed significantly higher than GPT-4 in its primary prediction (35 % vs. 0 %, *P* = 0.0083), but did not differ significantly at its top five (45 % vs. 25 %, *P* = 0.3203, ns) and top ten prediction (60 % vs. 30 %, *P* = 0.1110, ns) for DIRA. In comparison, the two did not significantly differ at their primary (60 % vs. 50 %, *P* = 0.7512, ns), top five (75 % vs. 65 %, *P* = 0.7311, ns), and top ten prediction (90 % vs. 65 %, *P* = 0.1274, ns) for FMF.

Compared to GPT-3.5, the internist performed significantly higher at its primary (35 % vs. 0 %, *P* = 0.0083), top five (45 % vs. 10 %, *P* = 0.0310), and top ten prediction (60 % vs. 16.67 %, *P* = 0.0046) for DIRA. For FMF, both did not differ significantly at its primary (60 % vs. 30 %, *P* = 0.1110, ns), top five (75 % vs. 50 %, *P* = 0.1908, ns), and top ten prediction (90 % vs. 60 %, *P* = 0.0648, ns).

The internist also performed significantly higher than LLaMa 2 at its primary (35 % vs. 0 %, *P* = 0.0083), top five (45 % vs. 0 %, *P* = 0.0012), and top ten prediction (60 % vs. 0 %, *P* < 0.0001) for DIRA. Compared to FMF, the internist also performed higher than LLaMa 2 at its primary (60 % vs. 0 %, *P* = 0004), top five (75 % vs. 10 %, *P* < 0.0001), and top ten prediction (90 % vs. 15 %, *P* < 0.0001).

When comparing GPT-4 to the GPT-3.5 model, there were no significant differences in accuracy in their primary prediction (50 % vs. 30 %, *P* = 0.3332, ns), top five (65 % vs. 50 %, *P* = 0.5231, ns), and top ten prediction (65 % vs. 60 %, *P* = >0.9999, ns) for FMF patients. For DIRA patients, there were no significant differences between the models in their primary (0 % vs. 0 %, *P* = >0.9999, ns), top five (25 % vs. 10 %, *P* = 0.4075, ns), and top ten prediction (30 % vs. 16.67 %, *P* = 0.4716, ns).

Moreover, comparing GPT-4 to LLaMa 2, there was a significant difference at its primary (50 % vs. 5 %, *P* = 0.0033), top five (65 % vs. 10 %, *P* = 0.0008), and top ten prediction (65 % vs. 15 %, *P* = 0.0031)
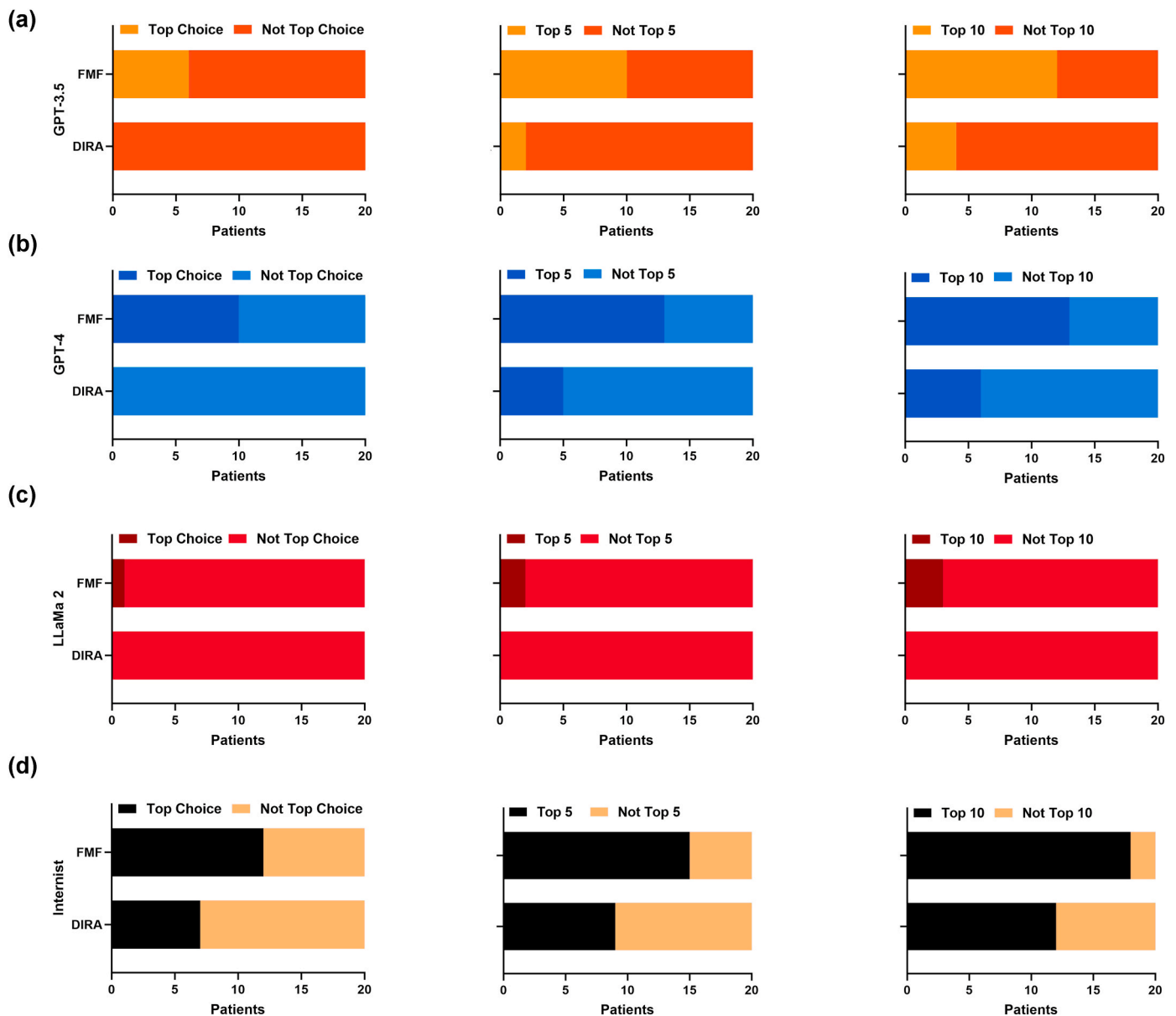
Fig. 1. (a) Diagnostic accuracy of GPT-3.5 at its primary choice (top choice), top five, and top ten diagnoses. (b) Diagnostic accuracy of GPT-4 at its primary choice (top choice), top five, and top ten diagnoses. (c) Diagnostic accuracy of LLaMa 2 at its primary choice (top choice), top five, and top ten diagnoses. (d) Diagnostic accuracy of the internist at its primary choice (top choice), top five, and top ten diagnoses.

for FMF. For DIRA, it was not significant at its primary prediction (0 % vs. 0 %, $P = > 0.9999$, ns). However, it was significant at its top five prediction (25 % vs. 0 %, $P = 0.0471$) and top ten prediction (30 % vs. 0 %, $P = 0.0202$).

Additionally, GPT-3.5 and LLaMa 2 did not differ significantly at its primary prediction (30 % vs. 5 %, $P = 0.0915$), but was significant at its top five (50 % vs. 10 %, $P = 0.0138$), and its top ten prediction (60 % vs. 15 %, $P = 0.0079$) for FMF. For DIRA, both models did not differ significantly at its primary (0 % vs. 0 %, $P = > 0.9999$, ns), top five (10 % vs. 0 %, $P = 0.4872$, ns), and top ten prediction (16.67 % vs. 0 %, $P = 0.1140$, ns).

All elements involved in the clinical vignette were statistically compared between both DIRA and FMF patients shown in Fig. 2. All values have been presented as the mean ± standard error of the mean. DIRA patients were 0.3084 ± 0.1506 years old, which was significantly younger compared to FMF patients that were 29.16 ± 3.863 years ($P < 0.0001$) shown in Fig. 2c. For ethnicity, DIRA patients were much more diverse (10 total) compared to FMF (5 total), shown in Fig. 2a–b. For sex,

there was no significant difference between both disease groups ($P = 0.4506$, ns). Although the majority of FMF vignettes did not have a reported consanguinity of patients, the majority of DIRA patients were non-consanguineous.

From the ultra-rarity of DIRA, there are no guidelines or official criteria for diagnosis. However, there is an official criteria for FMF reported by Livneh et al. [44]. We have listed all of the major and minor criterias from the clinical presentation of all vignettes in Table 1 and Table 2, respectively.

4. Discussion

In this study, we evaluated the diagnostic accuracy of the GPT-4, GPT-3.5, and LLaMa 2 models in 40 clinical vignettes of DIRA and FMF compared to an internal medicine physician.

Given the overall results of all cases, there were no AI models that predicted DIRA as its primary prediction. All of the correct predictions were either in the top five or top ten diagnoses. The symptoms of DIRA
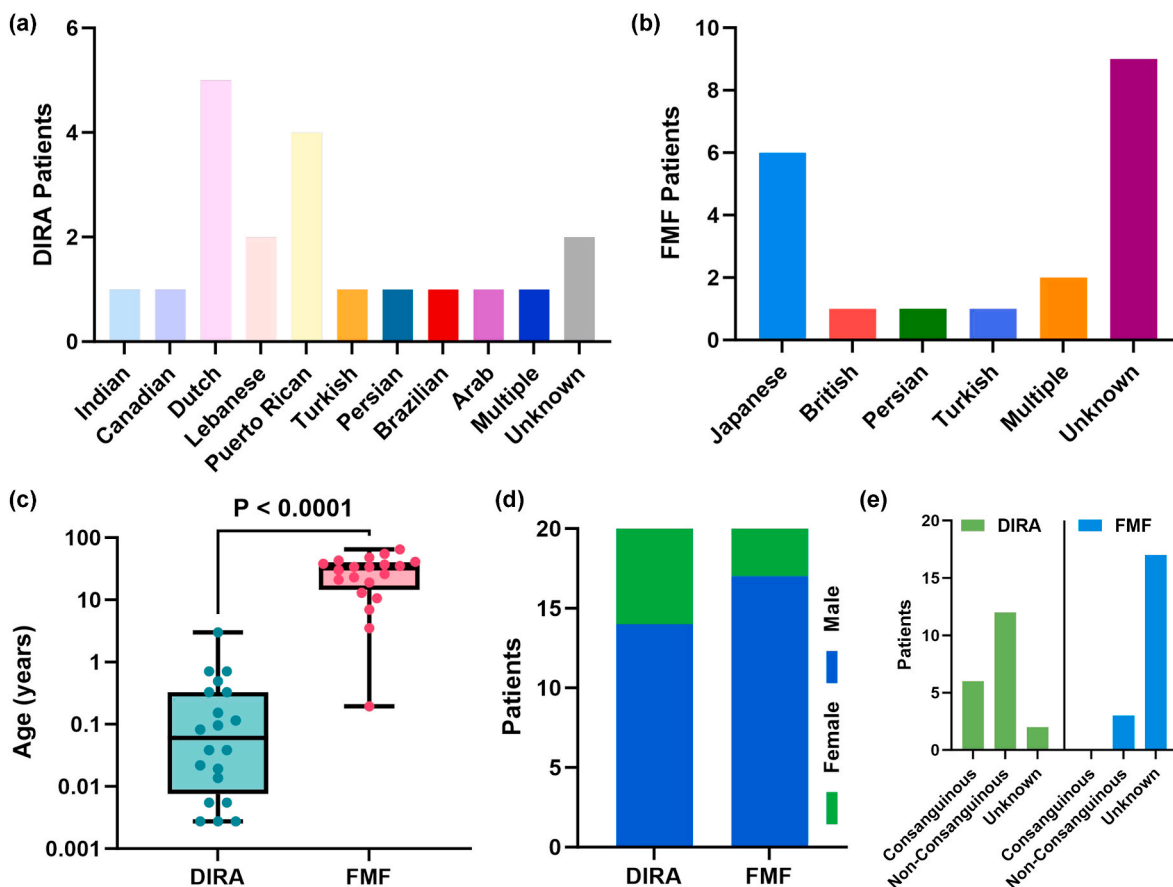
Fig. 2. (a) Ethnicities of DIRA patients. (b) Ethnicities of FMF patients. (c) Age of DIRA patients vs. Age of FMF patients. (d) Sex of DIRA and FMF patients. (e) Consanguinity of DIRA and FMF patients.

Table 1
Major criteria or typical attacks from FMF in all clinical vignettes. Simplified criteria model obtained from Bashardoust et al. [47].

Major Criteria	$n = 20$	Percentage (%)
Generalized Peronstitis	0	0
Pericarditis or Unilateral Pleuritis	2	10
Fever Alone	19	95
Incomplete Abdominal Attack	16	80
Monoarthritis (hip, knee, ankle)	1	5

Table 2
Minor criteria or incomplete attacks from FMF in all clinical vignettes. Simplified criteria model obtained from Bashardoust et al. [47].

Minor Criteria	$n = 20$	Percentage (%)
Chest	6	30
Joint	4	20
Favorable Response to Colchicine	1	5
Exertional Leg Pain	0	0

suggest similarities to Chronic Recurrent Multifocal Osteomyelitis and Neonatal-Onset Multisystem Inflammatory Disease. Given that DIRA is a monogenic form of CRMO, the overlapping similarities with DIRA may be leading to a lower certainty of DIRA within AI models [45]. At the same time, NOMID has similar symptoms to DIRA such as systemic inflammation, skin rashes, and joint pain, which could also be leading to incorrect diagnoses by the model [46]. Without thorough genetic screening, it would be a difficult task for LLMs to differentiate rare autoinflammatory disorders such as DIRA from other similar disorders.

For FMF, the major elements of clinical presentation were abdominal pain and fever and the minor elements were chest and joint pain. These unique symptoms assisted in achieving a higher diagnostic accuracy of all FMF patients compared to DIRA patients for GPT-4 and GPT-3.5. Even though GPT-4 had a lower accuracy than the internist, the difference remained not significant. With additional training data and tuning of hyperparameters adjusted to the unique elements of FMF, AI models could potentially have higher accuracy in diagnosis.

However, this study has multiple limitations. Due to the rarity of DIRA, there are a low number of reported research articles and case reports for it. This reduced the sample size of the study to 20 vignettes per each disease. This study demonstrated the capability of AI models in challenging rheumatological diseases without any genetic screening. A higher sample of AI models and physicians will be investigated in the future.

To the best of our knowledge, this is the first study that evaluates the diagnostic accuracy of LLMs in autoinflammatory disorders, specifically DIRA and FMF. GPT-4 was able to correctly identify up to 65 % of FMF patients, not significantly different from an internist with 90 % identification. For DIRA, GPT-4 was able to correctly identify 30 % of patients, while the internist at 60 % identification. However, the internist performed significantly higher in both diseases compared to GPT-3.5 and LLaMa 2. Given the preliminary metrics of LLMs for their diagnostic accuracy, further testing is required to determine the accuracy of models in more difficult and similar rheumatological diseases.

Funding

Author Contributions

J.P.P. - Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Visualization, and Writing - original draft. K.P.P. - Project Administration, Supervision, Writing - original draft, and Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Not applicable.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jtauto.2023.100213.

References

[1] Introducing ChatGPT, OpenAI. https://openai.com/blog/chatgpt, 2022.
[2] Introducing LlaMa, A Foundational, 65-Billion-Parameter Large Language Model, Meta AI, 2023. https://ai.meta.com/blog/large-language-model-llama-meta-ai/.
[3] X. Yang, A. Chen, N. PourNejatian, H.C. Shin, K.E. Smith, C. Parisien, C.B. Compas, C. Martin, A. Costa, M.G. Flores, Y. Zhang, T. Magoc, C.A. Harle, G. Lipori, D. A. Mitchell, W.R. Hogan, L. Shenkman, J. Bian, Y. Wu, A large language model for electronic health records, Npj Digital Medicine 5 (1) (2022), https://doi.org/10.1038/s41746-022-00742-2.
[4] R. Qureshi, D.T. Shaughnessy, K.a.R Gill, K.A. Robinson, T. Li, E. Agai, Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? Syst. Rev. 12 (1) (2023) https://doi.org/10.1186/s13643-023-02243-z.
[5] T.H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo, V. Tseng, Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models, PLOS Digital Health 2 (2) (2023), e0000198, https://doi.org/10.1371/journal.pdig.0000198.
[6] J. Kottlors, G. Bratke, P. Rauen, C. Kabbasch, T. Persigehl, M. Schlamann, S. Lennartz, Feasibility of differential diagnosis based on imaging patterns using a large language model, Radiology 308 (1) (2023), https://doi.org/10.1148/radiol.231167.
[7] Z. Kanjee, B. Crowe, A. Rodman, Accuracy of a generative artificial intelligence model in a complex diagnostic challenge, JAMA 330 (1) (2023) 78, https://doi.org/10.1001/jama.2023.8288.
[8] T. Hirosawa, Y. Harada, M. Yokose, T. Sakamoto, R. Kawamura, T. Shimizu, Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study, Int. J. Environ. Res. Publ. Health 20 (4) (2023) 3378, https://doi.org/10.3390/ijerph20043378.
[9] P. Galozzi, L. Punzi, P. Sfriso, Clinical overlapping in autoinflammatory diseases: the role of gene duplication, Front. Immunol. 8 (2017), https://doi.org/10.3389/fimmu.2017.00392.
[10] S.T. Reddy, J. Song, R. Geoffrey, R. Lorier, M. Suchi, U. Broeckel, M.J. Hessner, J. W. Verbsky, An autoinflammatory disease due to homozygous deletion of the *IL1RN* locus, N. Engl. J. Med. 360 (23) (2009) 2438–2444, https://doi.org/10.1056/nejmoa0809568.
[11] F. Onen, Familial mediterranean fever, Rheumatol. Int. 26 (6) (2005) 489–496, https://doi.org/10.1007/s00296-005-0074-3.
[12] L.O. Mendonça, L. Malle, F.X. Donovan, S.C. Chandrasekharappa, G.a.M Sanchez, M. Garg, U. Tedgård, M. Castells, S.S. Saini, S. Dutta, R. Goldbach-Mansky, D. Suri, A.A. De Jesus, Deficiency of interleukin-1 receptor Antagonist (DIRA): report of the first Indian patient and a novel deletion affecting IL1RN, J. Clin. Immunol. 37 (5) (2017) 445–451, https://doi.org/10.1007/s10875-017-0399-1.
[13] I. Aksentijevich, S.L. Masters, P.J. Ferguson, P. Dancey, J. Frenkel, A. Van Royen-Kerkhoff, R.M. Laxer, U. Tedgård, E.W. Cowen, T. Pham, M.G. Booty, J.D. Estes, N. G. Sandler, N. Plass, D.L. Stone, M.L. Turner, S. Hill, J.A. Butman, R. Schneider, R. Goldbach-Mansky, An autoinflammatory disease with deficiency of the interleukin-1–receptor antagonist, N. Engl. J. Med. 360 (23) (2009) 2426–2437, https://doi.org/10.1056/nejmoa0807865.

[14] B. Sözeri, B. Gerçeker-Türk, B. Yıldız-Atıkan, S. Mir, A. Berdeli, A novel mutation of interleukin-1 receptor antagonist (il1rn) in a dira patient from Turkey: diagnosis and treatment, Turk. J. Pediatr. 60 (5) (2018) 588, https://doi.org/10.24953/turkjped.2018.05.020.
[15] J. Kuemmerle-Deschner, T. Welzel, K. Hoertnagel, I. Tsiflikas, A. Hospach, X. Liu, S. Schlipf, S. Hansmann, S.D. Samba, A. Griesinger, S.M. Benseler, A.N. Weber, New variant in the IL1RN-gene (DIRA) associated with late-onset, CRMO-like presentation, Rheumatology 59 (11) (2020) 3259–3263, https://doi.org/10.1093/rheumatology/keaa119.
[16] W. Sakran, S.A. Shalev, H. El-Shanti, Y. Uziel, Chronic recurrent multifocal osteomyelitis and deficiency of interleukin-1–receptor antagonist, Pediatr. Infect. Dis. J. 32 (1) (2013) 94, https://doi.org/10.1097/inf.0b013e3182700cc1.
[17] C. Schnellbacher, G. Ciocca, R. Menendez, I. Aksentijevich, R. Goldbach-Mansky, A.M. Duarte, R. Rivas-Chacon, Deficiency of interleukin-1 receptor antagonist responsive to anakinra, Pediatr. Dermatol. 30 (6) (2012) 758–760, https://doi.org/10.1111/j.1525-1470.2012.01725.x.
[18] V. Ziaee, L. Youssefian, M. Faghankhani, A. Jazayeri, A.H. Saeidian, H. Vahidnezhad, J. Uitto, Homozygous IL1RN mutation in siblings with deficiency of interleukin-1 receptor antagonist (DIRA), J. Clin. Immunol. 40 (4) (2020) 637–642, https://doi.org/10.1007/s10875-020-00767-w.
[19] C.N. Brau-Javier, J. Gonzales-Chavez, J.R. Toro, Chronic cutaneous pustulosis due to a 175-kb deletion on chromosome 2q13, Arch. Dermatol. 148 (3) (2012) 301, https://doi.org/10.1001/archdermatol.2011.2857.
[20] P.G. Thacker, L.A. Binkovitz, K.B. Thomas, Deficiency of interleukin-1-receptor antagonist syndrome: a rare auto-inflammatory condition that mimics multiple classic radiographic findings, Pediatr. Radiol. 42 (4) (2011) 495–498, https://doi.org/10.1007/s00247-011-2208-y.
[21] L.O. Mendonça, A. Grossi, F. Caroli, R.A. De Oliveira, J. Kalil, F.F.M. Castro, A. Pontillo, I. Ceccherini, M.M.S.B. De Barros, M. Gattorno, A case report of a novel compound heterozygous mutation in a Brazilian patient with deficiency of Interleukin-1 receptor antagonist (DIRA), Pediatr. Rheumatol. 18 (1) (2020), https://doi.org/10.1186/s12969-020-00454-5.
[22] R. Abdwani, E. Abdalla, B.A. Masilhi, A. Shalaby, A. Al-Maawali, Novel mutation in interleukin 1 receptor antagonist associated with chronic diarrhoea in infancy, J. Paediatr. Child Health 58 (1) (2022) 186–188, https://doi.org/10.1111/jpc.15440.
[23] K. Minkis, I. Aksentijevich, R. Goldbach-Mansky, C.M. Magro, R.A. Scott, J. G. Davis, N. Sardana, R. Herzog, Interleukin 1 receptor antagonist deficiency presenting as infantile pustulosis mimicking infantile pustular psoriasis, Arch. Dermatol. 148 (6) (2012), https://doi.org/10.1001/archdermatol.2011.3208.
[24] Y. Hotta, T. Kawasaki, T. Kotani, H. Okada, K. Ikeda, S. Yamane, N. Yamada, S. Sekoguchi, Y. Isozaki, Y. Nagao, M. Murotani, H. Oyamada, Familial mediterranean fever without fever, Intern. Med. 59 (10) (2020) 1267–1270, https://doi.org/10.2169/internalmedicine.3175-19.
[25] N. Cliff-Patel, B. Yusuf, S. Hamdani, V. Ziauddin, Familial Mediterranean fever: a differential diagnosis for the surgical abdomen, JRSM Open 13 (9) (2022), 205427042211234, https://doi.org/10.1177/20542704221123433.
[26] H. Ariga, R. Kunisaki, T. Ojima, S. Suzuki, K. Okada, J. Kashimura, Familial Mediterranean fever with colonic lesions: a case report, DEN Open 4 (1) (2023), https://doi.org/10.1002/deo2.246.
[27] A. Kübra, S.U. Sarici, G. Kolukisa, D. Altun, Familial mediterranean fever with neonatal onset: case report, Case Reports in Pediatrics (2020) 1–3, https://doi.org/10.1155/2020/6649525, 2020.
[28] M.C. Maggio, M.S. Castiglia, G. Corsello, Familial Mediterranean Fever: an unusual cause of liver disease, Ital. J. Pediatr. 45 (1) (2019), https://doi.org/10.1186/s13052-019-0712-0.
[29] K. Kasamaki, C. Kusano, H. Ikehara, S. Suzuki, M. Esaki, A. Irie, K. Hayashi, H. Okuno, M. Moriyama, T. Gotoda, Familial mediterranean fever with small bowel stenosis, Internal Medicine (2019), https://doi.org/10.2169/internalmedicine.2293-18.
[30] N. Evans, J. Ray, C.M. Prather, Atypical Autosomal-Dominant inheritance of familial Mediterranean fever, ACG Case Reports J. 8 (3) (2021), e00525, https://doi.org/10.14309/crj.0000000000000525.
[31] S. Matsumoto, S. Urayoshi, Y. Yoshida, Familial Mediterranean fever in which Crohn's disease was suspected: a case report, BMC Res. Notes 7 (1) (2014), https://doi.org/10.1186/1756-0500-7-678.
[32] M. Takahashi, T. Ebe, T. Kohara, M. Inagaki, H. Isonuma, I. Hibiya, T. Mori, K. Watanabe, H. Ikemoto, Periodic fever compatible with familial mediterranean fever, Intern. Med. 31 (7) (1992) 893–898, https://doi.org/10.2169/internalmedicine.31.893.
[33] A. Malik, J. Malik, J. Javaid, H.S. Khan, M. Mohsin, M. Shoaib, A case of aortic dissection in familial Mediterranean fever, J. Cardiol. Cases 24 (6) (2021) 296–299, https://doi.org/10.1016/j.jccase.2021.05.005.
[34] K. Ashida, E. Terada, A. Nagayama, S. Sakamoto, N. Hasuzawa, M. Nomura, Necessity of utilizing physiological glucocorticoids for managing familial Mediterranean fever, Am. J. Case Reports (2020), https://doi.org/10.12659/ajcr.920983.
[35] M. Toshida, Y. Konishi, A. Ikenouchi, N. Okamoto, R. Yoshimura, Colchicine-Resistant Familial Mediterranean Fever with Depressive State Successfully Treated with Escitalopram, Cureus (2021), https://doi.org/10.7759/cureus.15145.
[36] A. Malek, T. Zeraati, A. Sadr-Nabavi, N. Vakili, M.R. Abbaszadegan, Cardiac tamponade: a rare manifestation of familial Mediterranean fever, Case Reports in Rheumatology (2022) 1–5, https://doi.org/10.1155/2022/8334375, 2022.
[37] T. Hosoi, K. Ishii, N. Tozaka, D. Kishida, Y. Sekijima, A. Tamaoka, Familial Mediterranean fever is important in the differential diagnosis of recurrent aseptic

meningitis in Japan, Intern. Med. 59 (1) (2020) 125–128, https://doi.org/10.2169/internalmedicine.3432-19.

[38] W.M. Darwish, S. Darwish, M. Darwish, B.F. Darwish, M680I/M694V heterozygous mutation in early onset familial Mediterranean fever, J. Med. Cases 12 (9) (2021) 351–354, https://doi.org/10.14740/jmc3747.

[39] Y.Ö. Atikel, B.E. Derinkuyu, S.A. Bakkaloglu, Unusual presentation of familial Mediterranean fever with co-existing polyarteritis nodosa and acute post-streptococcal glomerulonephritis, Clinical Case Reports 10 (7) (2022), https://doi.org/10.1002/ccr3.6022.

[40] K. Asakura, S. Yanai, S. Nakamura, K. Kawaski, M. Eizuka, K. Ishida, M. Endo, T. Sugai, K. Migita, T. Matsumoto, Familial Mediterranean fever mimicking Crohn disease, Medicine 97 (1) (2018) e9547, https://doi.org/10.1097/md.0000000000009547.

[41] Ş. Erten, C. Erzurum, A. Altunoglu, Three family members with familial mediterranean fever carrying the M694V mutation showed different clinical presentations, Intern. Med. 51 (13) (2012) 1765–1768, https://doi.org/10.2169/internalmedicine.51.7537.

[42] K. Iwata, T. Toma, A. Yachie, Atypical familial mediterranean fever presenting with recurrent migratory polyarthritis, Intern. Med. (2019), https://doi.org/10.2169/internalmedicine.3001-19.

[43] Y. Lin, M. Zhu, Z. Su, The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials, Contemp. Clin. Trials 45 (2015) 21–25, https://doi.org/10.1016/j.cct.2015.07.011.

[44] A. Livneh, P. Langevitz, D. Zemer, N. Zaks, S. Kees, T. Lidar, A. Migdal, S. Padeh, M. Pras, Criteria for the diagnosis of familial mediterranean fever, Arthritis Rheum. 40 (10) (1997) 1879–1885, https://doi.org/10.1002/art.1780401023.

[45] A.P. Rao, P.P. Mallya, S. Ranjani, J. Raghuram, Chronic recurrent multifocal osteomyelitis - a case series from India, Indian J. Orthop. 52 (6) (2018) 672–677, https://doi.org/10.4103/ortho.ijortho_464_17.

[46] R. Goldbach-Mansky, Current status of understanding the pathogenesis and management of patients with NOMID/CINCA, Curr. Rheumatol. Rep. 13 (2) (2011) 123–131, https://doi.org/10.1007/s11926-011-0165-y.

[47] B. Bashardoust, Familial Mediterranean fever; diagnosis, treatment, and complications, PubMed 4 (1) (2015) 5–8. https://pubmed.ncbi.nlm.nih.gov/28197466.