Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 273–280 DOI: 10.15439/2023F1683 ISSN 2300-5963 ACSIS, Vol. 37

Interactive, Personalized Decision Support in Analyzing Women's Menstrual Disorders

Łukasz Sosnowski
(0000-0003-2388-4008)
Systems Research Institute,
Polish Academy of Sciences
Newelska 6, Warsaw, Poland
sosnowsl@ibspan.waw.pl

Soma Dutta
(0000-0002-7670-3154)
University of Warmia and Mazury in Olsztyn
Słoneczna 54, Olsztyn, Poland
soma.dutta@matman.uwm.edu.pl

Iwona Szymusik
(0000-0001-8106-5428)
Dep. of Obstetrics, Perinatology and Neonatology,
Center of Postgraduate Medical Education,
Ceglowska 80, Warsaw, Poland
iwona.szymusik@gmail.com

Wojciech Chaber
OvuFriend Sp. z o.o.,
Zlota 61/100, 00-819 Warsaw, Poland
wojciech.chaber@ovufriend.com

Paulina Kasprowicz
OvuFriend Sp. z o.o.,
Zlota 61/100, 00-819 Warsaw, Poland
paulina.kasprowicz@ovufriend.com

***Abstract*—This paper is in continuation to the paper published in FedCSIS 2022. In the earlier paper we presented the general scheme behind the AI based model for determining the possible ovulation dates as well as the possibility of some health risks. Here apart from the already discussed schemes for Premenstrual Syndrome (PMS), Luteal Phase Defect (LPD), and polyp and fibroids, a few additional schemes like hypothyroidism, polycystic ovary syndrom (PCOS) are included. Moreover, we attempt to throw light on the novelty of this AI based scheme from the perspective personalized, case sensitive, interactive medical support which does not depend only on a preset rule based system for diagnosing diseases.**

***Index Terms*—Medical decision support, Interactive AI, Explainable AI**

I. Introduction

FROM *the emergence of Artificial Intelligence, many researchers from different aspects contributed towards a broader goal of an artificially intelligent agent. However the progress is still far from the level of reaching close to human-like reasoning. This may be the reason that the current literature on AI showcases examples where the researchers specify the need by introducing terms like 'human-centered AI' [1], 'human-in-the-loop of machine learning' [2] etc.* explainable Reflection of similar thoughts can be noticed also in the context of decision support for different health care systems, such as IBM's dream project Watson, which was supposed to revolutionize everything from diagnosing patients and recommending treatment options to finding candidates for clinical trials; however, it failed as instead of trained with real data it was trained with hypothetical cases provided by a small group of doctors. The reason is quite understandable as expecting that the model's success with test data will directly translate to the real world does not really meet in reality. So, one way to improve the performance of an AI system is to engage users in providing feedback in order to continually improve the model. Thus, the terms like personalized medicine [3], evidence based medicine [4] are becoming prevalent in the literature of AI.

The main feature in both [3], [4] is to create such protocols for medical care that combine the knowledge from the existing literature of medicine, experience of the professionals, as well as the input parameters, habits, life style, and preferences of the individual patients. Moreover, in order to be sure that such data are properly gathered as well as to guide a patient during the intermediate steps of performing tests required for the diagnosis, an interactive interface among different stakeholders of the AI system is also required. So, it is quite clear that such a paradigm combines together different kinds of physical entities, information associated to them (e.g., a patient and her informational base, a team of experts and their informational bases) and their interactions, which as a whole indicates a real physical process of computation on a complex granule [5], [6].

In continuation to a series of papers [7]–[10], here we present the developments made in the platform of OvuFriend[1] focusing on introducing the above mentioned aspects in an AI system for helping women in determining the possibility of conceiving and understanding the hidden risk of health problems based on their input. The platform of OvuFriend 1.0 was developed as a part of R&D project where through a mobile app an user can put the data related to her physical and mental states during a specific menstrual cycle, and the underlying algorithm of the app helps to get an analysis of the possibility of conceiving or not conceiving. The second stage of OvuFriend's project, namely OvuFriend 2.0, focuses on the

Co-financed by the EU Smart Growth Operational Program 2014-2020 under the project "Developing innovative solutions in the domain of detection of frequent intimate and hormonal health disorders in women of procreative age based on artificial intelligence and machine learning - OvuFriend 2.0", POIR.01.01.01-00-0826/20.

[1] www.ovufriend.pl

analysis of whether a particular user has the possibility of having the risk of Premenstrual Syndrome (PMS) Luteal Phase Defect (LPD), benign growths like polyps, fibroids in the uterus, Polycystic Ovary Syndrome (PCOS) or hypothyroidism Among them in [11] a detailed discussion regarding the schemes of PMS, LPD, and the risk from polyp and fibroids are discussed. Here we add the schemes corresponding to hypothyroidism and PCOS.

Apart from discussing the two new schemes, namely hypothyroidism and PCOS, in this paper one of our main targets is to present in what sense the AI model of OvuFriend incorporates the features like (i) learning and updating based on real data (ii) adaptation of diagnosing strategies based on interactions with users and medical as well as analytical experts, and (iii) visual as well as linguistic explainability of the relationship between the gathered data and their labelling. Inclusion of these features makes the OvuFriend platform for women's healthcare more close to the above mentioned AI paradigm of interactive, personalized, evidence based healthcare support systems keeping human in the loop.

The paper is organized as follows. Section II presents a brief general description of the scheme running behind the OvuFriend app as well as the schemes analyzing certain health risks based on a complete cycle data of a user. The schemes for hypothyroidism and PCOS, requiring a sequence of cycles data, are presented in Section III. Section IV presents the process of building the reference set based on which the app can decide effectively over new cases. In particular we would emphasize on the novelty of the process of selecting reference set which allows a team based interactive environment among the experts, the user and the consultant in the process of deciding how, when and for whom which strategy of treating and diagnosing should be selected so that the data gathered from them can be used in the reference set with certain reliability. Section V presents concluding remarks.

II. General Scheme of Ovulation and health risks

The general scheme in OvuFriend 2.0 for having an AI based app determining the possible days of ovulation as well as the possibility of different health risks is developed based on three hierarchical levels, known as *Detector level*, *Cycle level*, and *User level*. In the detector level the user can put information related to her mental and physical health over (at least) one complete cycle. Relative to the need a set of attributes is set by the medical experts. Based on the input of a particular user the values for those attributes are determined by a team of medical experts. From the values of the attributes from a completed cycle, the cycle level concepts such as *ovulation happened*, *days of ovulation*, *follicular phase interval*, *luteal phase interval*, *PMS score* etc are determined. In the user level the system aggregates the data related to the detector level as well as the cycle level concepts of a particular user for a finitely many cycles. *Risk of PMS*, *risk of LPD*, *risk of infertility* etc are a few examples of the user level concepts. For a cycle level concept, the system calculates the probabilistic ratio of the concerned cycle level concept over the total number of cycles considered for a particular user. Moreover, the system is also fed with a threshold value for each such concept. The threshold value for a particular concept is learned and with time this is updated based on the opinions of the medical experts and the histories of already recorded and analysed cases. If the respective ratio for a particular user level concept is greater than the prefixed threshold for that concept the user is notified about the possibility of such health risk.

As prerequisite the data related to the physical and mental health of a woman before, during, and after a complete menstrual cycle is collected. After gathering data over a complete cycle (or a few consecutive cycles) the analysis for different health risks starts. Initially, the data is processed to investigate whether the ovulation has occurred and then based on that to find the possible days of its occurrence. At this stage all the detector level concepts are analysed. If through the primary analysis it is determinned that ovulation has been occurred, then an attempt is made to indicate two intervals of equal length falling into the follicular phase and the luteal phase of the concerned cycle respectively [11].

A. Summary of schemes requiring one complete cycle data

The schemes requiring a complete cycle data, namely PMS, LPD, polyp and fibroids, are already discussed in [11]. Though the basic formulas for calculating these health risks are different, the general form of the underlying algorithms is similar.

To analyze the risk of PMS, which is a combination of symptoms that many women get about a week or two before their period, the coefficients of occurrence of the physical symptoms and mood symptoms are calculated (see [11]). The set of symptoms and formulas for calculating the coefficients based on them are defined with the help of a team of medical experts. From the user's input all physical and mood symptoms are counted for both the phases P_1 and P_2. Aggregating the number of physical and mood symptoms in a phase the coefficients are calculated according to the following formulas.

$$P_iMoodFeelCoeff = \frac{(SumOfOccurrenceP_iMood)}{K_1 \times PhaseLength} \times \alpha + (1-\alpha) \tag{1}$$

where i = 1, 2 and $\alpha \in (0,1)$,

$$P_2PhysFeelCoeff = \frac{(SumOfOccurrenceP_2Phys)}{K_2 \times PhaseLength} \times \beta + (1-\beta) \tag{2}$$

where $\beta \in (0,1)$.

The symbols $SumOfOccurrenceP_iMood$ and $SumOfOccurrenceP_iPhys$ respectively indicate the number of mood and the number of physical symptoms occurred in a particular phase P_i, and K_1 and K_2 represent respectively the total number of all moods and physical symptoms listed in the system. The factors α and β represent the significance of the given components in the respective coefficients. Using the above coefficients *PMS score*, denoted as PMS_{score}, is calculated by the following formula.

$$PMS_{score} = \frac{P_2MoodFeelCoeff}{P_1MoodFeelCoeff} + \frac{P_2PhysFeelCoeff}{w_1} \tag{3}$$

where w_1 is the weight chosen by a team of medical experts.

Luteal Phase Defect (LPD) is a health condition that may play a role in infertility. The general prerequisite for determining the risk of LPD [12] is same as what is discussed above. The specific formula that is fed to the algorithm in order to calculate the susceptibility of LPD (Equation 4) is as follows.

$$LPD_{score} = w_1 * LutParameters + w_2 * DecFer \tag{4}$$

The parameters $LutParameters, DecFer \in [0, 1]$ respectively denote the values for *Luteal Phase Parameters* and *Decreased Fertility*. The *Luteal Phase Parameters* are determined based on the luteal phase length and various other factors related to the analysis of bleeding during the luteal phase. The *Decreased Fertility* depends on the period of time in which the attempts are made for conceiving a child, the number of miscarriages etc. The values for $LutParameters, DecFer$ are obtained based on the input data of a particular user, and w_1, w_2 are some weights that are chosen by the team of experts based on their collective knowledge regarding the significance of $LutParameters$ and $DecFer$ in indicating LPD.

Presence of fibroids and polyps too may cause infertility and recurrent pregnancy loss. The algorithm starts with checking whether ovulation has occurred. The primary analysis focuses on the data related to inter-menstrual bleeding or spots. Then examining the cycle level concepts and associated symptoms characterizing polyp or fibroids starts. The values for disordered menstruation ($DisMens$), decreased fertility ($DecFer$), and the values for physical symptoms related to such diseases ($PhysSymp$) are obtained from the input data of the user. Then the following score is calculated.

$$Score = w_1 * DisMens + w_2 * DecFer + w_3 * PhysSymp \tag{5}$$

The weights w_1, w_2, w_3 are chosen by the team of experts. All these values are scaled in the interval $[0, 1]$ based on the information related to inter-menstrual bleeding, long-lasting menstruation, intensity of menstruation, miscarriage, long trying time for conceiving, pelvis pain, polyuria etc.

In each of the above contexts, for a given user the grade of the susceptibility of a particular disease is calculated by considering $\frac{k}{n}$ if in k such cycles, out of n cycles, the susceptibility of the respective disease is detected.

III. Schemes requiring consecutive cycles' data

In this section we present two newly analyzed health risks, namyly PCOS and hypothyroidism, which require a sequence of consecutive cycles' data of a user.

A. Scheme for PCOS

PCOS creates a condition where the ovaries produce an abnormal amount of androgens, that are usually present in women in small amounts [13]. Contrary to the above mentioned schemes, to analyse the risk of PCOS the algorithm needs the data of the user for a few months. Based on the detector level parameters such as stress, appetite, depression, hypersensitivity, insomnia, problem in concentration, BMI, length of cycle etc., relevant cycle level concepts such as *increasing level of anxiety*, *lower self-esteem*, *family history of PCOS*, *long cycle*, *extended trying time for conceiving* etc are determined. Some of the above mentioned cycle level concepts are marked with binary values and some with fuzzy values, on a scale of $0 \leq 0.33 \leq 0.66 \leq 1$; these values are marked over a span of time. After completion of a cycle, all the relevant cycle level concepts are determined. Each of the considered cycles is then characterized with the help of these concepts described on a multidimensional time series.

Some groups of symptoms are analyzed by qualitative as well as quantitative indicators. For example, it is checked whether any of the symptoms belonging to the group occurred at least once on a given day (qualitative), as well as how many symptoms (quantitative) from the group occurred on a day. The frequency of occurrence of a symptom usually is analyzed based on the selected time period. For instance, the occurrence of the symptom 'fatigue' 4 times in a 45-day cycle may indicate the greater possibility for anxiety than occurrence of the same symptom 4 times during half of the time span of the cycle. Compare to the above schemes here the algorithm chooses the next plan of actions based on an interaction with the user. There are different forms available for deeper analysis of some of the above mentioned detector or cycle level concepts. If a user meets the PCOS boundary conditions, she is asked to provide some specific parameters in the follicular phase of the cycle for consecutive 3 days. If the user rates them three times negatively, the label for *low self esteem* is activated. Then further the user is led to complete a more detailed low self-esteem survey.

The analysis for PCOS also starts with checking the possibility of ovulation and determining respective intervals. To enable PCOS susceptibility analysis, the input for the cycle must be completed for at least 10 days; the same data for previous two cycles must also be available meeting the same conditions. Symptoms for PCOS persist for a long time. So, one cycle may not reliably assess the presence of PCOS. Moreover, exploring three consecutive cycles increases the likelihood of the observations of the user. For each of these series of cycles, possible ovulation is determined.

The coefficient $cycle_nScore$ for the nth cycle is calculated based on the following formula.

$$\begin{aligned} cycle_nScore = X_{1n} * w_1 + X_{2n} * w_2 + X_{3n} * w_3 + \\ X_{4n} * w_4 + X_{5n} * w_5 + X_{6n} * w_6 + X_{7n} * w_7 \end{aligned} \tag{6}$$

where X_{in} is calculated based on the number points obtained for the i-th group of concepts that have appeared in the n-th cycle. For example, $X_{5n} = \frac{increased\_anxiety+depressive\_mood}{2}$ indicates that the two operands in the numerator represent the number of points obtained for those two parameters from the 5-th group of concepts in the n-th cycle. The weights w_i, $1 \leq i \leq n$ are selected based on the significance of a group of symptoms over other. Then the sum of the points of each cycle from the sequence is added and normalized according to the formula below.

$$normScore = \frac{\Sigma_{i=1}^{3} cycle_iScore}{3 * \Sigma_{j=1}^{7} w_j} \tag{7}$$

Based on the values for $nomScore$, different possible sequences of cycles, recorded over a time period, are ranked in the descending order. The two chosen sequences of three cycles can be such that one of the cycles can be the first in one sequence and middle in another sequence. So, the sequence with highest $normScore$ is selected for the analysis of PCOS.

The scheme of PCOS is presented in Fig. 1. To determine the PCOS susceptibility one sequence of cycles, which is completed in last six months, is selected from the history of a user. If among a series of cycles at least two are detected with a vulnerability of PCOS, the respective user is assigned to PCOS risk. Then, at user level the degree of risk is calculated based on the ratio of the number of PCOS-susceptible cycles to the number of months over which the observation is made.

B. Scheme for Hypothyroidism

In hypothyroidism the thyroid gland does not produce enough thyroid hormones, leading to changes in the menstrual cycle. The scheme for hypothyroidism is quite similar to the scheme for PCOS. Here also the algorithm requires data for three consecutive cycles. Data for all the cycles are processed to test determine the date of onset of ovulation as well as the detector and the cycle level concepts. If, in each of the cycles from the sequence, enough data is marked for the algorithm to determine the occurrence of ovulation, the algorithm proceeds to the next stage. The cycle level concepts and the symptoms, such as feeling cold, feeling sleepy, concentration problems, decreased appetite, constipation, swelling, decreased libido, memory problems, etc., which are relevant to hypothyroidism, are selected. Then for each cycle a score, denoted as Sc_n, is determined from the sequence using the following formula.

$$\begin{aligned} Sc_n &= w_1 * PhySym_n + w_2 * Len_n + w_3 * Ov_n \\ &\quad + w_4 * DecFer_n + w_5 * PsySym_n \end{aligned} \tag{8}$$

The weights $w_1, \ldots, w_5$ are chosen by the experts, and the values of the parameters are computed from the input of user. The symbol $PhySym_n$ denotes the value corresponding to the physical symptoms during the n-th cycle, Len_n indicates the length of the n-th cycle, Ov_n corresponds to the number of ovulations occurred in the n-th cycle, $DecFer_n$ stands for the value of the decreased fertility in the n-th cycle, and $PsySym_n$ represents the value corresponding to the psychological symptoms in the n-th cycle. From the scores of three consecutive cycles the score for the risk of hypothyroidism is calculated for the whole sequence using the following formula.

$$Score_{Hypth} = Sc_1 + Sc_2 + Sc_3 \tag{9}$$

where 1, 2, 3 denote the numbers of the cycles in the sequence.

If the score is greater than or equal to the preset threshold, the most recent cycle in the sequence is assigned a hypothyroidism susceptibility at the cycle level and the score is then calculated just by adding the score obtained in three consecutive cycles. The score obtained for each such single cycle from a chain of three consecutive cycles is used to assess the risk of developing hypothyroidism at the level of the user. In this process all completed cycles, that have occurred during the last n months, are selected. Then all possible sequence combinations of three consecutive cycles are created from them, and the sum of the scores is calculated for each sequence. If the sum of the scores for any of the sequences is greater than or equal to the pre-fixed cut-off value, a risk of hypothyroidism is assigned to the user, and a grade is calculated in the range of $[0, 1]$. After the analysis of a user's risk for hypothyroidism the data and analysis, obtained from the sequences, are again assessed by medical experts. Based on such history of sequences the cut-off point is updated.

IV. Interactively adapting treatment and diagnosis strategy based on users' perception

We now attempt to illustrate the key features of OvuFriend's application which allow to create an interface for telemedicine consultation and choose appropriate course of actions based on analyzed data of a user. Through the interactive interface a user, a team of experts (medical and analytic), and a consultant together may share a platform for interacting with queries and respective answers, uploading and scanning documents/results, presenting an illustrative graphical representation of causes and outcomes related to a concept, and choosing labels for certain values of parameters based on consensus. In this regard, we present the design of some components and their roles contributing towards the working strategy of the app.

A. Building reference set incorporating real data through interactions

In Introduction we have discussed about failure of different decision support systems trained based on hypothetical data. Here, the reference set, for training the model of the app, has been chosen from three different populations of users.

One population pertains to the already registered users of the app for whom certain vulnerabilities are detected on the basis of physical symptoms, mood symptoms and parameters of menstrual cycles declared in the system. Based on the data recorded in the cycles of the users further medical examinations are suggested. Then based on context, indicated by a precise flowchart of the algorithm, the users are selected to be included in the reference set when some specific results are confirmed by blood tests, TSH, Testosterone, progesterone, ultrasound examination of the reproductive organs etc., or on the basis of a questionnaire completed in the app, serving as a medical teleconsultation. The second population pertains to women who have participated in a questionnaire survey conducted on the OvuFriend's platform and have declared certain diseases voluntarily. For such users based on the results of survey uploaded to the system they are selected for inclusion in the reference set. The third population pertains to the women who as a part of marketing activities of the app are envouraged to declare problems of having certain diseases or noticing symptoms from a given set of relevant symptoms on the OvuFriend's platform. In response to their willingness to take part in the project, they are offered free medical examinations, and in case of positive result for some

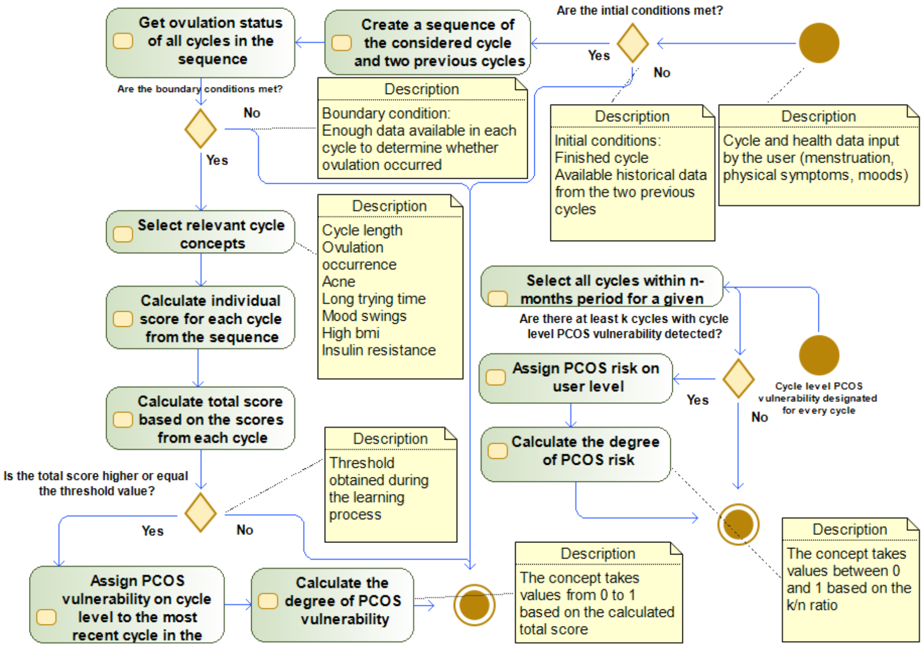

Get ovulation status of all cycles in the sequence
Create a sequence of the considered cycle and two previous cycles
Are the intial conditions met?
Yes
No
Description
Boundary condition:
Enough data available in each cycle to determine whether ovulation occurred
Are the boundary conditions met?
No
Yes
Description
Initial conditions:
Finished cycle
Available historical data from the two previous cycles
Description
Cycle and health data input by the user (menstruation, physical symptoms, moods)
Select relevant cycle concepts
Description
Cycle length
Ovulation occurrence
Acne
Long trying time
Mood swings
High bmi
Insulin resistance
Calculate individual score for each cycle from the sequence
Calculate total score based on the scores from each cycle
Select all cycles within n-months period for a given
Are there at least k cycles with cycle level PCOS vulnerability detected?
Assign PCOS risk on user level
Yes
No
Cycle level PCOS vulnerability designated for every cycle
Calculate the degree of PCOS risk
Description
Threshold obtained during the learning process
Is the total score higher or equal the threshold value?
Yes
No
Assign PCOS vulnerability on cycle level to the most recent cycle in the
Calculate the degree of PCOS vulnerability
Description
The concept takes values from 0 to 1 based on the calculated total score
Description
The concept takes values between 0 and 1 based on the k/n ratio

Fig. 1. Complete scheme for determining risk vulnerability of PCOS (2 diagrams)

tests they are given access to the Ovufriend's application to upload their results. Then further data on historical cycles are collected during a survey conducted by a consultant.

The consultant contacts using the data provided by the application or the OvuFriend's platform. If some conditions are met the consultant asks questions about the symptoms, relative to the analyzed disease. The questionnaire is designed by a medical expert. So, though all the questionnaire surveys and teleconsultation processes are conducted within the algorithmic set up of the app, it includes both human in the loop as well as real physical interactions.

Moreover, the users after completion of the tests the scans of medical examinations upload to the system or sent to OvuFriend's platform. The exchange of information between the consultant and the team of experts is carried out using spreadsheets saved on a cloud drive, due to which it is possible to track the editions by all team members. The information obtained by the user, including medical tests, and the answers given during surveys, are then checked by the analytical team, in cooperation with a medical expert.

B. Explainable model storing and labeling reference set data

An explainable AI model is another great challenge on which the present days AI development is still struggling. OvuFriend's model is capable to address the above mentioned challenge to some extent. In particular, it refers to the part of the model where each user's data along with the scans of test results are stored against an uniquely generated user-id.

The data obtained in the survey along with the test results are uploaded by the consultant in the cloud environment for review of the analytical team. On a regular interval all the uploaded information of the users are analyzed by the analytical team and a medical expert. If based on the initial information a user is considered suitable for the project, a user-id is created and the information details is saved in the database. Fig. 2 shows a fragment of a relational database, presenting directory containing files with records of medical examinations of different project users.

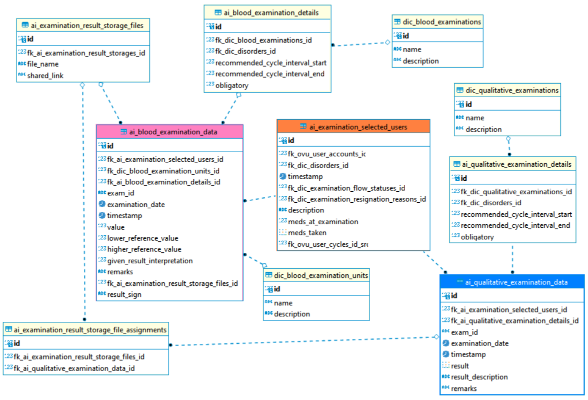

ai_examination_result_storage_files
id
fk_ai_examination_result_storages_id
file_name
shared_link
ai_blood_examination_details
id
fk_dic_blood_examinations_id
fk_dic_disorders_id
recommended_cycle_interval_start
recommended_cycle_interval_end
obligatory
dic_blood_examinations
id
name
description
dic_qualitative_examinations
id
name
description
ai_blood_examination_data
id
fk_ai_examination_selected_users_id
fk_dic_blood_examination_units_id
fk_ai_blood_examination_details_id
exam_id
examination_date
timestamp
value
lower_reference_value
higher_reference_value
given_result_interpretation
remarks
fk_ai_examination_result_storage_files_id
result_sign
ai_examination_selected_users
id
fk_ovu_user_accounts_ic
fk_dic_disorders_id
timestamp
fk_dic_examination_flow_statuses_id
fk_dic_examination_resignation_reasons_id
description
meds_at_examination
meds_taken
fk_ovu_user_cycles_id_src
ai_qualitative_examination_details
id
fk_dic_qualitative_examinations_id
fk_dic_disorders_id
recommended_cycle_interval_start
recommended_cycle_interval_end
obligatory
dic_blood_examination_units
id
name
description
ai_qualitative_examination_data
id
fk_ai_examination_selected_users_id
fk_ai_qualitative_examination_details_id
exam_id
examination_date
timestamp
result
result_description
remarks
ai_examination_result_storage_file_assignments
id
fk_ai_examination_result_storage_files_id
fk_ai_qualitative_examination_data_id

Fig. 2. Relational database for storing information related to a user and creating links for navigating between relevant information

Medical examinations stored in the database are presented to the doctor by means of a visualization that combines the information provided by the users, in the form of scans of tests and the data registered in the app for individual cycles. More specifically, the visual representation related to the measurement details and symptoms of a user also includes additional information obtained from the database, e.g. cycle type, cycle status, registered drugs, anomaly detector result and descriptive information regarding the analysis of the expert tagged with a comment and information obtained from the user during the consultation before performing tests. The visualizations in this project are based on the experience gathered in previous projects based on a well-known approach from other areas using fuzzy linguistic summaries widely described in [14], [15]. This type of visualization allows for a comprehensive presentation of all archival information registered by the user in the app including information from the labeling stage, through the period in which the test is performed, to the current cycle at the time of presentation of the final results. At the next stage based on the visualizations presented to the team of experts the selection of the final labeling of the reference set is performed.

The model of creating such visualization serves two aspects of explainability. In one hand, it presents a comprehensive visual representation of the data with all notes and comments from the user and the medical expert, and on the other hand it presents a visual relationship between the results of the analysis made by the app and the measurement data registered in the app depending on particular disease. For each of the anomalies, i.e, thyroid diseases, PCOS, NFL etc., the respective visual representation is created during consultations with a medical expert. That is, based on the data saved on cloud against each user-id, the medical expert can create some cause-effect relations among the measurement data and the concerned diseases, and that information gets translated to the system creating a visual representation of the selected relations using some software packages for time series analysis.

Fig. 3 presents how through a spreadsheet visualization of all data relative to a particular patient is presented in a compact and comprehensive way to the analytical team as well as to the user. In Fig. 3 the information presents values of different parameters over three consecutive cycles, length of each of which is presented in the header. In the left hand side using a sliding option for going up and down one can check information concerning a particular day over this sequence of cycles, and in the bottom the labels are chosen automatically by the algorithm based on values of the parameters entered from user's input.

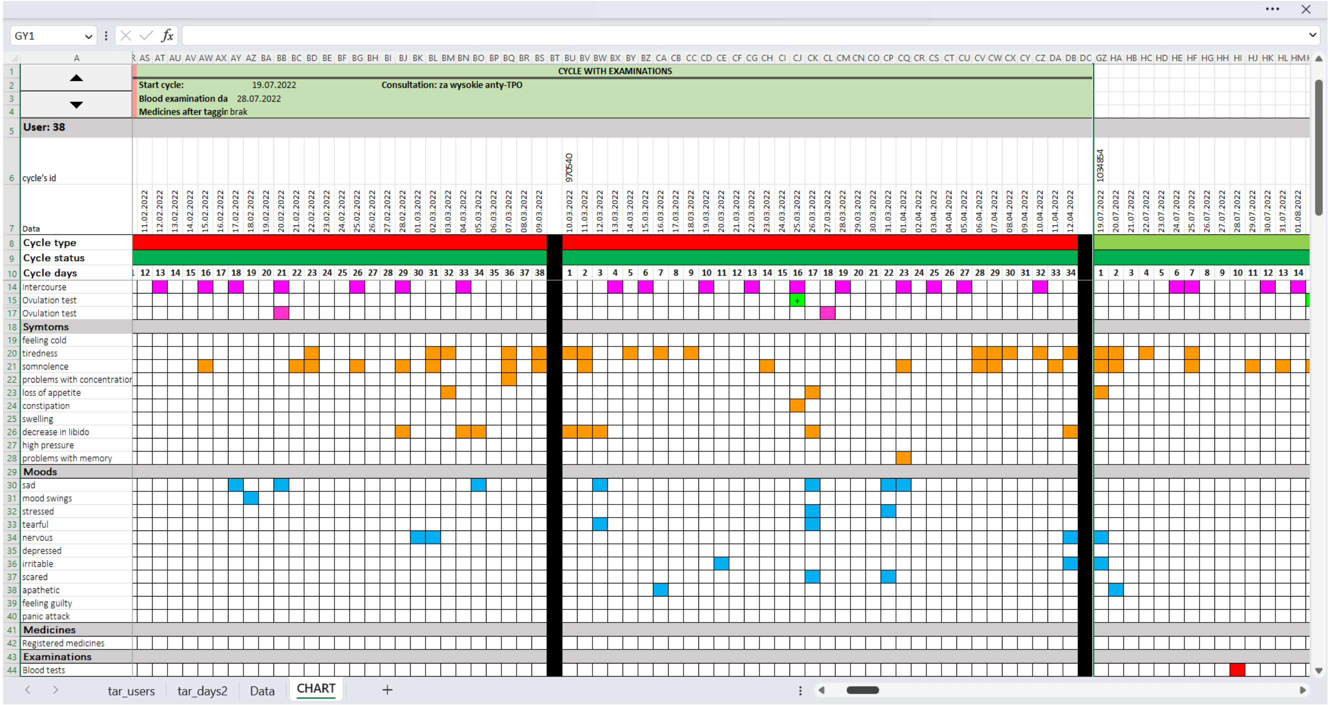
The visualization for each disease consists of two files.

(i) One is a sql file in which data is generated for each user included in the app. Here the data presents a user and her cycles divided into days. The range of days selected for visualization depends on the number of cycles recorded in the app and varies depending on the number of cycles entered and their lengths for each user. A rule is fed to the app to create the visualization; the time axis is defined by the initial data related to one cycle or three cycles used in the labeling process. The cycle, in which the tests are performed, are marked as the anchor points. From the tagged cycles, a maximum of three cycles are searched back. From here the data is supposed to be represented by visualization. All the cycles (or cycle) included in the app for labeling, including the cycle in which tests have been performed, are presented. As the cycle, in which tests are performed, is marked on the chart, the performed transformations saved in this file lead to two main tables: the users table and the days table. The user table contains information such as: chart number, user_id, information about the tagged cycle (cycle_id, length, start date), information about tagging by the expert (shipped package number, order in the package, expert tagging result, comment), information about medications taken, date of the test, test result, link to the test file, type of test, comment obtained after contacting the user etc. The days table contains data for visualized cycles for each user in a specific package, including: user_id, cycle_id, cycle order on the chart, date, cycle day, information about mucus, cervix, bleeding, intercourse, ovulation tests, pregnancy, symptoms, moods, concepts, as well as detector indications, cycle type and information about cycle status.

(ii) The other one is a xlsx file in which data prepared with the use of SQL code are read. Using the ODBC connection to the PostgreSQL database, previously prepared tables with users and days are uploaded (saved in the .sql file). Next, the data is transformed in order to visualize them on the timeline, the length of which varies based on the number of cycles registered by the user in the app. The tab with the chart shows the graphical form of the automatically transformed data, depending on the refresh of the data in the .sql file, by defining the appropriate package number. Switching between the users is possible using the user selection control in the form of arrows, a vertical slider scrolling between visualization sections and a horizontal slider scrolling between user cycles.

The presence of information in a line is conditioned by its color, depending on the day of occurrence. The vertical black bars are used to separate the cycles from each other. In the right of a black marker a new cycle starts with counting of the days in the cycle and the intensity of bleeding over days is represented on the graph. Cycles are presented on a timeline from the oldest to the newest. The users whose cycles already have been labeled by a medical expert are selected for the visualization of cycles after the tests; that is, sequentially first they participate in the tagging process, receive a referral for tests from a given medical package depending on the disease, perform the test and send the scans of the results, which are saved in the appropriate folder in the google drive, which can be accessed by the team of experts and the consultant responsible for contacting the users. The user-ids, corresponding to the selected cycles for presentation, are fed to the packages by which visualizations are created.

The visualizations are presented in such a way that the medical expert can have an insight into the widest possible range of information of the patients. The whole presentation is realized in an interactive way. A medical expert, using the buttons in the upper left corner of each of the presented visualizations, can switch tabs and obtain different information related to a chosen user. On the other hand, using the horizontal scroll bar



CYCLE WITH EXAMINATIONS

Start cycle: 19.07.2022 Consultation: za wysokie anty-TPO

Blood examination da 28.07.2022

Medicines after taggin brak

User: 38

cycle's id: 970540, 1034854

Fig. 3. Hypothyroidism labelling form prepared for medical experts to evaluate susceptibility of selected cycles. Form is based on such attributes as: bleeding, mucus, bbt, cervix, intercourse, feeling cold, tiredness, somnolence, sad, mood swings, tearful, stressed, nervous, depressed, irritable, scared, problem with the concentration, sleep disturbance, constipation, etc.

it is possible to navigate from one cycle to other and obtain a view of the complete history of the user's recorded cycles.

V. Experimental Results: Concluding Remarks

In Section IV, we presented the design of reference set as one of the unique selling points of OvuFriend's application. In this section we would present a brief summary of the experimental results obtained based on the chosen reference set. In contrary to the experimental results obtained in the earlier stages [11], here we present the experimental results based on the actual users whose health risks or anomalies have been analyzed by OvuFriend's schemes.

The reference set consists of a list of users assigned to the selected anomaly with the actual class specified by the physician. Subsets for individual anomalies are balanced in terms of the number of positive and negative classes, so as to obtain a similar number of elements in both the classes. As the different methods of data processing depend on the anomaly, the users have been grouped by anomaly, not by a group of diseases. Later the final evaluation is calculated based on the average results of evaluations performed for different anomalies falling within a group of diseases. For example, in case of LPD 57 cases from each of positive and negative classes are selected; while in case of PCOS the reference set contains 94 cases from each of positive and negative classes.

For evaluating the effectiveness of each of the algorithms four experiments have been conducted for each of the disorders. Using ReSample evaluation [16] each of the experiments is conducted such as 1000, 500, 100, and 10 times respectively. Two disjoint subsets are designated as the training set and test set where the former contains 33% of the reference set and the later consists of remaining 67%. Evolutionary algorithms with a fitting function based on a combination of the accuracy measure and the F1Score measure are used to train the respective thresholds for all the disorders and these values are learned on each iteration of the training set containing 33% of the tagged cycles sample in particular disorders. The test procedure is performed on remaining cycles in given disorders which accounted for 67%. For each iteration results are stored in the contingency table. Then all True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are calculated to find the effectiveness measures of the algorithms. Due to page limitation here only the result for 1000 repetitions is presented (cf. Table I). presents .

The label TP means that cycle is tagged with at least 0.6 by the medical experts and is classified as a positive case of the disease by the algorithm; whereas, the case for FP is determined when the experts have given mark below 0.6 but the algorithm has classified the case into positive class. The case for TN is obtained when the experts have assigned less than 0.6 and the algorithm has classified as negative as well. Finally the cases for FN is indicated when the algorithm has classified as negative but the expert evaluated as greater or equal 0.6.

From Table I it is visible that the obtained results are quite satisfactory; especially comparing to the experimental

TABLE I
Results averaged over 1000 iterations of the ReSample routine. Abbreviations: # - sample, TP - True Positives, TN - True Negatives, FP - False Positives, FN - False Negatives, PR - Precision, RE - Recall, F1 - F1 score, $_{mn}$- min, $_{mx}$- max, AC - accuracy, PCOS - Polycystic Ovary Syndrome, HYP - hypothyroidism, LPD - Luteal phase deficiency

Type	Sample	TP	TN	FP	FN	PR	RE	F1	AC	F1_mn	F1_mx	AC_mn	AC_mx
HYP	126000	61472	39922	23096	1510	0.727	0.976	0.833	0.805	0.757	0.902	0.730	0.881
PMS	68000	37448	13668	7783	9101	0.828	0.804	0.816	0.752	0.575	0.905	0.544	0.868
PCOS	126000	60643	42814	20065	2478	0.751	0.961	0.843	0.821	0.358	0.903	0.516	0.889
LPD	76000	35526	24555	13342	2577	0.727	0.932	0.817	0.791	0.089	0.899	0.461	0.882

results obtained in the earlier stage [11] based on hypothetical data, here the experimental results based on real data is commendably good. Moreover, apart from the quantitative values showing satisfactory experimental results, in this paper, our main emphasize has been on the qualitative worth of the AI model which attempts to address a few important at the same time challenging aspects of AI. Precisely the novelty of the approach includes building of an AI model which is (i) trained on real data, (ii) sensitive to user's perceptions, (iii) able to learn and revise through interactions among the stakeholders (such as a user and the (medical) experts), (iv) designed to adapt suitable strategies for the required course of actions (e.g., suggesting further tests or filling up additional questionnaire etc.) in the process of decision making, and (v) possessing the ability of explainability of its decision to the stakeholders through a process of visualization. As a whole the proposed model is a good attempt towards the goals envisaged by the paradigms like personalized, evidence based medicine [3], [4], human-centered AI [1], and IGrC [5], [6].

However, there are a few aspects where the model has limitations. Firstly, it is difficult to collect a very large collection reference data as use of the application and participation in the project is voluntary. Moreover, usually users, who are trying to get pregnant, are interested in using the app; whereas for reference data only historical data that was recorded before pregnancy can be used. Another limitation is related to full implementation of the developed algorithms and preparing them to work in a real environment. Simultaneously in many places replacing manual process of setting parameters and weights (by experts) by ML, so that all relevant parameters are learned from reference data sets, is also required.

References

[1] Ben Shneiderman. *Human – Centered AI.* Oxford University Press, Oxford, UK, 2022.

[2] Robert (Munro) Monarch. *Human-in-the-Loop Machine Learning. Active Learning and Annotation for Human-Centered AI.* MANNING, Shelter Island, NY, 2021.

[3] Carlos Fernández-Llatas and Jorge Munoz-Gama *et al*. *"Process Mining in Healthcare"*, pages 41–52. Springer, 2020.

[4] Margaret A. Hamburg and Francis S. Collins. "The Path to Personalized Medicine". *New England Journal of Medicine*, 363(4):301–304, 2010.

[5] Soma Dutta and Andrzej Skowron. Interactive granular computing model for intelligent systems. In Z. Shi, M. Chakraborty, and S. Kar, editors, *Intelligence Science III. 4th IFIP TC 12 International Conference (ICIS 2020), Durgapur, India, February 24-27, 2021, Revised Selected Papers*, volume 623 of *IFIP Advances in Information and Communication Technology (IFIPAICT) book series*, pages 37–48. Springer, Cham, Switzerland, 2021.

[6] Soma Dutta and Andrzej Skowron. Interactive Granular Computing Connecting Abstract and Physical Worlds: An Example. In Holger Schlingloff and Thomas Vogel, editors, *Proceedings of the 29th International Workshop on Concurrency, Specification and Programming (CS&P 2021), Berlin, Germany, September 27-28, 2021*, volume 2951 of *CEUR Workshop Proceedings*, pages 46–59. CEUR-WS.org, 2021.

[7] Lukasz Sosnowski and Tomasz Penza. "Generating Fuzzy Linguistic Summaries for Menstrual Cycles". volume 21 of *Annals of Computer Science and Information Systems*, pages 119–128, 2020.

[8] Joanna Fedorowicz, Lukasz Sosnowski, Dominik Slezak, Iwona Szymusik, Wojciech Chaber, Lukasz Milobedzki, Tomasz Penza, Jadwiga Sosnowska, Katarzyna Wójcicka, and Karol Zaleski. "Multivariate Ovulation Window Detection at OvuFriend". In Tamás Mihálydeák, Fan Min, Guoyin Wang, Mohua Banerjee, Ivo Düntsch, Zbigniew Suraj, and Davide Ciucci, editors, *Rough Sets - International Joint Conference, IJCRS 2019, Debrecen, Hungary, June 17-21, 2019, Proceedings*, volume 11499 of *Lecture Notes in Computer Science*, pages 395–408. Springer, 2019.

[9] Lukasz Sosnowski, Iwona Szymusik, and Tomasz Penza. "Network of Fuzzy Comparators for Ovulation Window Prediction". volume 1239 of *Communications in Computer and Information Science*, pages 800–813. Springer, 2020.

[10] Lukasz Sosnowski and Jakub Wróblewski. "Toward automatic assessment of a risk of women's health disorders based on ontology decision models and menstrual cycle analysis". In Yixin Chen, Heiko Ludwig, Yicheng Tu, Usama M. Fayyad, Xingquan Zhu, Xiaohua Hu, Suren Byna, Xiong Liu, Jianping Zhang, Shirui Pan, Vagelis Papalexakis, Jianwu Wang, Alfredo Cuzzocrea, and Carlos Ordonez, editors, *2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, December 15-18, 2021*, pages 5544–5552. IEEE, 2021.

[11] Lukasz Sosnowski, Joanna Zulawinska, Soma Dutta, Iwona Szymusik, Aleksandra Zygula, and Elzbieta Bambul-Mazurek. Artificial intelligence in personalized healthcare analysis for womens' menstrual health disorders. In Maria Ganzha, Leszek A. Maciaszek, Marcin Paprzycki, and Dominik Slezak, editors, *Proceedings of the 17th Conference on Computer Science and Intelligence Systems, FedCSIS 2022, Sofia, Bulgaria, September 4-7, 2022*, volume 30 of *Annals of Computer Science and Information Systems*, pages 751–760, 2022.

[12] Kenneth A. Ginsburg. "Luteal Phase Defect: Etiology, Diagnosis, and Management". *Endocrinology and Metabolism Clinics of North America*, 21(1):85–104, 1992. Reproductive Endocrinology.

[13] Neil F. Goodman, Rhoda H. Cobin, Walter Futterweit, Jennifer S. Glueck, Richard S. Legro, and Enrico Carmina. "American Association of Clinical Endocrinologists, American College of Endocrinology, and Androgen Excess and PCOS Society Disease State Clinical Review: Guide to the Best Practices in the Evaluation and Treatment of Polycystic Ovary Syndrome - Part 1". *Endocrine Practice*, 21(11):1291–1300, 2015.

[14] Janusz Kacprzyk and Ronald R. Yager. "Linguistic summaries of data using fuzzy logic". *International Journal of General Systems*, 30(2):133–154, 2001.

[15] Janusz Kacprzyk and Slawomir Zadrozny. "Fuzzy logic-based linguistic summaries of time series: a powerful tool for discovering knowledge on time varying processes and systems under imprecision". *Wiley Interdiscip. Rev. Date Min Knowl. Discov.*, 6(1):37–46, 2016.

[16] P.I. Good. *"Resampling Methods: A Practical Guide to Data Analysis"*. Birkhäuser Boston, 2005.