Development and clinical validation of real-time artificial intelligence diagnostic companion for fetal ultrasound examination

J. J. STIRNEMANN[1,2], R. BESSON[3], E. SPAGGIARI[1,2,4], S. ROJO[3], F. LOGE[3], H. PEYRO-SAINT-PAUL[3], S. ALLASSONNIERE[5,6], E. LE PENNEC[6,7], C. HUTCHINSON[8], N. SEBIRE[8] and Y. VILLE[1,2]

[1]*Department of Obstetrics and Maternal–Fetal Medicine, Necker–Enfants Malades Hospital, AP-HP, Paris, France;* [2]*EA7328 Université de Paris, IMAGINE Institute, Paris, France;* [3]*SONIO SAS, Paris, France;* [4]*Department of Histology–Embryology and Cytogenetics, Unit of Embryo and Fetal Pathology, Necker–Enfants Malades Hospital, AP-HP, Paris, France;* [5]*School of Medicine, Université de Paris, INRIA EPI HEKA, INSERM UMR 1138, Sorbonne Université, Paris, France;* [6]*Center for Applied Mathematics, Ecole Polytechnique, Institut Polytechnique de Paris, Paris, France;* [7]*Xpop, INRIA Saclay Center, Paris, France;* [8]*NIHR Great Ormond Street Hospital Biomedical Research Centre, London, UK*

CONTRIBUTION

What are the novel findings of this work?
In this validation study, an artificial intelligence (AI)-based diagnostic support system showed high performance for the identification of fetal syndromes on ultrasound.

What are the clinical implications of this work?
AI-enabled software could bring complex and overlooked knowledge to care-providers involved in prenatal diagnosis. Expected benefits include: improving the diagnosis of rare fetal conditions; providing a standardized examination in case of an anomaly; limiting unnecessary repeat examinations and parental anxiety; and offering continuous medical education.

ABSTRACT

Objective Prenatal diagnosis of a rare disease on ultrasound relies on a physician's ability to remember an intractable amount of knowledge. We developed a real-time decision support system (DSS) that suggests, at each step of the examination, the next phenotypic feature to assess, optimizing the diagnostic pathway to the smallest number of possible diagnoses. The objective of this study was to evaluate the performance of this real-time DSS using clinical data.

Methods This validation study was conducted on a database of 549 perinatal phenotypes collected from two referral centers (one in France and one in the UK). Inclusion criteria were: at least one anomaly was visible on fetal ultrasound after 11 weeks' gestation; the anomaly was confirmed postnatally; an associated rare disease was confirmed or ruled out based on postnatal/postmortem investigation, including physical examination, genetic testing and imaging; and, when confirmed, the syndrome was known by the DSS software. The cases were assessed retrospectively by the software, using either the full phenotype as a single input, or a stepwise input of phenotypic features, as prompted by the software, mimicking its use in a real-life clinical setting. Adjudication of discordant cases, in which there was disagreement between the DSS output and the postnatally confirmed ('ascertained') diagnosis, was performed by a panel of external experts. The proportion of ascertained diagnoses within the software's top-10 differential diagnoses output was evaluated, as well as the sensitivity and specificity of the software to select correctly as its best guess a syndromic or isolated condition.

Results The dataset covered 110/408 (27%) diagnoses within the software's database, yielding a cumulative prevalence of 83%. For syndromic cases, the ascertained diagnosis was within the top-10 list in 93% and 83% of cases using the full-phenotype and stepwise input, respectively, after adjudication. The full-phenotype and stepwise approaches were associated, respectively, with a specificity of 94% and 96% and a sensitivity of 99% and

ORIGINAL PAPER

84%. *The stepwise approach required an average of 13 queries to reach the final set of diagnoses.*

Conclusions The DSS showed high performance when applied to real-world data. This validation study suggests that such software can improve perinatal care, efficiently providing complex and otherwise overlooked knowledge to care-providers involved in ultrasound-based prenatal diagnosis. © 2023 The Authors. *Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of International Society of Ultrasound in Obstetrics and Gynecology.*

INTRODUCTION

A fetal malformation or anomaly is detected on prenatal ultrasound examination in around 2–5% of all pregnancies[1]. Irrespective of its severity, the diagnosis of an anomaly raises the question of whether it is an isolated finding or if it is associated with other anomalies within a chromosomal or genetic syndrome that weighs negatively on fetal prognosis. There are over 9000 rare diseases in the Orphanet database[2], of which a few hundred express a consistent phenotype identifiable on prenatal ultrasound. However, the diversity of these syndromes and of many of their constitutive, often non-specific, features is beyond the knowledge of most specialists in prenatal diagnosis. Concurrently, ultrasound technology and technicity, driven by growing demand, have reached high levels of detail, raising the expectations of pregnant women for the assessment of fetal development.

We have developed a real-time decision support system (DSS)[3] that operates by suggesting, at each step of the ultrasound examination, the best phenotypic feature to assess next, in order to optimize the diagnostic pathway to the smallest number of possible diagnoses[4]. This DSS is knowledge-based[5] and comprises a dedicated database constructed from several sources and two dedicated algorithms. This assistant was tested on a large database of fetal anomalies.

METHODS

Description of software

The DSS software is named Sonio Expert; a video showing its software features is available online[6]. It was built on the association of two algorithms and a tailored database of prenatal syndromes and diseases with their respective ultrasonographic anomalies.

Database of anomalies and syndromes

The main sources for the database of fetal diseases are Orphanet[2] and the Centre de référence des agents tératogènes (CRAT)[7]. Orphanet is an open-source database of rare disorders maintained by an international consortium. The CRAT is the French teratogenicity reference center, maintained by the Department of Pharmacology at Trousseau Hospital in Paris. These databases were reviewed by a panel of experts from the fetal medicine unit at Necker Hospital in Paris, yielding a final database of 408 diseases with prenatal onset, visible symptoms on fetal ultrasound and a prevalence of between 1/400 (Down syndrome) and 'very rare' (i.e. for which the prevalence is unknown and only case reports exist). The database of ultrasound phenotypic anomalies was built using the Human Phenotype Ontology (HPO), which provides standardized terminology and a tree structure[8,9]. Some anomalies are 'ascendants' of others, meaning that they describe a more general concept (for example, 'abnormal morphology of the heart' is an ascendant of 'tetralogy of Fallot').

Software algorithm

The software was built on a combination of two algorithms: one for the diagnosis of a syndrome given the phenotype, and the second to suggest the next ultrasonographic anomaly for which to look[10]. The workflow starts from a 'call' anomaly or a risk factor, and displays at each iteration a number of phenotypic features for the physician to assess within the anatomical area currently being explored, contextual questions regarding possible risk factors and the current probability of the most likely diagnoses. The process ends by providing a list of differential diagnoses ordered by probability; the first is Sonio Expert's best guess, and the top-10 list presents the 10 most likely diagnoses.

The first algorithm performs Bayes' formula, fitted to work with medical ontologies[8,11] and their tree structure. The algorithm can deal with causal links between anomalies, some of which, such as Pierre Robin sequence, are complex entities that include other anomalies, such as cleft palate or microretrognathia. The algorithm is also able to process contextual information, such as fetal sex and gestational age, that change the probability of the syndromes.

The second algorithm is a decision tree based on average information gain[12,13]. Several refinements were necessary to produce acceptable diagnostic pathways for the user. First, we accounted for the ergonomics of the ultrasound examination, limiting back-and-forth movement between anatomical regions. The software also performs causal reasoning first, before considering an association of phenotypic anomalies as independent features of a syndrome.

Design and setting of validation study

The performance of the software was assessed on postmortem and postnatal data from the databases of two centers, Necker Hospital Fetal Medicine Department (Paris, France) and Great Ormond Street Hospital (London, UK), from 2007 to 2022. Eligible cases were those with at least one visible anomaly on fetal ultrasound after 11 weeks' gestation, with postnatal confirmation and with complete postnatal/postmortem investigation (physical examination, genetic testing and imaging). The database was curated and structured using HPO nomenclature.

A total of 713 cases were assessed (657 postmortem and 56 postnatal). Two cases were excluded because the

syndrome was not within Sonio's database, and 162 cases were excluded because a final diagnosis was not reached following diagnostic investigation and the condition could not be considered as isolated (i.e. the presence of a syndrome could not be ruled out). Therefore, the database used for the validation of the software comprised 549 cases for which an 'ascertained diagnosis' was available; the presence of a syndrome (constitutive of Sonio's database) was either confirmed or ruled out. This dataset therefore contains: (i) all syndromic cases for which a diagnosis was ascertained based on genetic findings (array comparative genomic hybridization, gene panel, exome sequencing) or phenotypic findings (X-ray, external, internal and histologic examinations) ($n = 317$); and (ii) cases for which investigations concluded that the anomaly was 'isolated', i.e. without a syndromic cause ($n = 232$). The phenotype of the latter cases includes the primary anomaly and its possible consequences (for example, polyhydramnios is a consequence of esophageal atresia).

This study was approved by the local ethics committees of Necker Hospital (CERPAHP.5#00011928) and Great Ormond Street Hospital (16/LO/0910).

Input models

The software was applied to each case using two models: 'full phenotype' and 'stepwise'. In the full-phenotype model, for each case all risk factors and observed anomalies are provided to Sonio Expert as a single input. This model provides an estimate of the crude performance of the system. In the stepwise model, the system is primed with a randomly selected 'typical' anomaly (i.e. with a probability of association with the syndrome of > 5%) as an input, and questions sequentially the presence or absence of specific anomalies. Permutation of the first input symptom yields several 'scenarios' for each case. In this model, the system only receives the information it has requested regarding the phenotype. An example of the application of both models to a real case is presented in Table 1.

Endpoints

Evaluation of the clinical performance of Sonio Expert was based on its sensitivity and specificity to differentiate a syndromic case from a non-syndromic case, using the full-phenotype and stepwise models. We also assessed the ability of the software to identify the ascertained diagnosis in its top-10 list as well as the number of steps needed to reach the final list of diagnoses using the stepwise model.

Management of discordance between Sonio Expert's output and ascertained diagnosis

A true positive was a case for which both the Sonio Expert's output and the ascertained diagnosis were

Table 1 Application of Sonio Expert using 'full-phenotype' and 'stepwise' models to a case of branchio-otorenal (BOR) syndrome, confirmed genetically, presenting with external ear malformation, renal hypoplasia, hydronephrosis, talipes and oligohydramnios, but without renal agenesis, renal dysplasia or preauricular skin tags

Input model	Top-5 differential diagnoses	Phenotypic anomalies recovered (n)
Full phenotype	1. BOR syndrome 2. Townes−Brocks syndrome 3. Branchio-oculo-facial syndrome 4. CHARGE syndrome 5. Distal trisomy 14q	NA
Stepwise starting from oligohydramnios	1. BOR syndrome 2. Townes−Brocks syndrome 3. Branchio-oculo-facial syndrome 4. CHARGE syndrome 5. Incidental association: external ear malformation/renal hypoplasia	3/5
Stepwise starting from hydronephrosis	1. BOR syndrome 2. Townes−Brocks syndrome 3. Distal trisomy 14q 4. Branchio-oculo-facial syndrome 5. CHARGE syndrome	3/5
Stepwise starting from renal hypoplasia	1. BOR syndrome 2. Townes−Brocks syndrome 3. Branchio-oculo-facial syndrome 4. Distal trisomy 14q 5. CHARGE syndrome	2/5
Stepwise starting from external ear malformation	1. Trisomy 18 2. Incidental association: talipes equinovarus/external ear malformation 3. Steinert myotonic dystrophy 4. BOR syndrome 5. Saethre−Chotzen syndrome	2/5

NA, not applicable.

concordant, i.e. the software's best guess was a syndrome and the ascertained diagnosis was within its top-10 list. Adjudication by a panel of external experts ensured that a final diagnosis was established for each case in which the software's output and ascertained diagnosis were discordant. The panel of six independent external experts included three expert geneticists (S.S., E.S., A.T.) and three fetal medicine experts (F.A., A.G., A.K.). Each discordant case was adjudicated by two randomly selected experts (always including one fetal medicine specialist and one geneticist). A third expert was asked to resolve disagreements between the two first experts. The experts were provided with the complete phenotype and blinded to the ascertained diagnosis and Sonio Expert's output. They were first asked if the case was suggestive of a syndrome or not, and second, which syndromes were suspected. The experts summarized their final decision as one of four options: (i) Sonio Expert and ascertained diagnosis agree (when the initial discordance was based on synonymous terms or on two subtypes of the same disease); (ii) the diagnosis was achievable given the phenotype and Sonio Expert failed; (iii) reaching the ascertained diagnosis was impossible given the ultrasound phenotype (i.e. the phenotype was non-specific or atypical); and (iv) the ascertained diagnosis was possibly incorrect.

RESULTS

The validation study was conducted on a database of 549 clinical cases with a final diagnosis, which covered 110 syndromes or diseases (88 single-gene, 15 chromosomal, three infectious and four toxic/teratogenic). Each syndromic case ($n = 317$) displayed a mean of 5.9 (range, 1–21) phenotypic anomalies, of which 2.8 were atypical of the disease (i.e. the anomaly was not reported or was reported as atypical in the sources used to build the database including Orphanet and the published literature). Each non-syndromic case ($n = 232$) displayed a mean of 1.4 (range, 1–5) phenotypic anomalies. The validation database covered 27% (110/408) of the syndromes present in Sonio Expert's database and 83% of their overall cumulative prevalence of fetal syndromes amenable to prenatal diagnosis by ultrasound. Chromosomal microarray analysis (CMA), karyotyping, fluorescence *in-situ* hybridization or genome sequencing was abnormal in at least 132/317 (42%) syndromic cases. The other diagnoses were based on clinical assessment.

Discordance between Sonio Expert's output and the ascertained diagnosis occurred in 65/317 syndromic cases, either because the best guess was an isolated abnormality (19/65; ascertained diagnosis was within top-10 list in 14/19 and not in top-10 in 5/19), or because the best guess was a syndrome but the ascertained diagnosis was not within the top-10 list (46/65) (Figure 1). These 65 cases presented with a mean of 5.7 phenotypic signs, of which 72% were atypical, according to Sonio Expert's database.

Discordance between Sonio Expert's output and the ascertained diagnosis occurred in 14 isolated non-syndromic cases. These 14 cases were considered false positives, were deemed to be failures and were not adjudicated upon (Table 2). Therefore, the specificity of the soft-
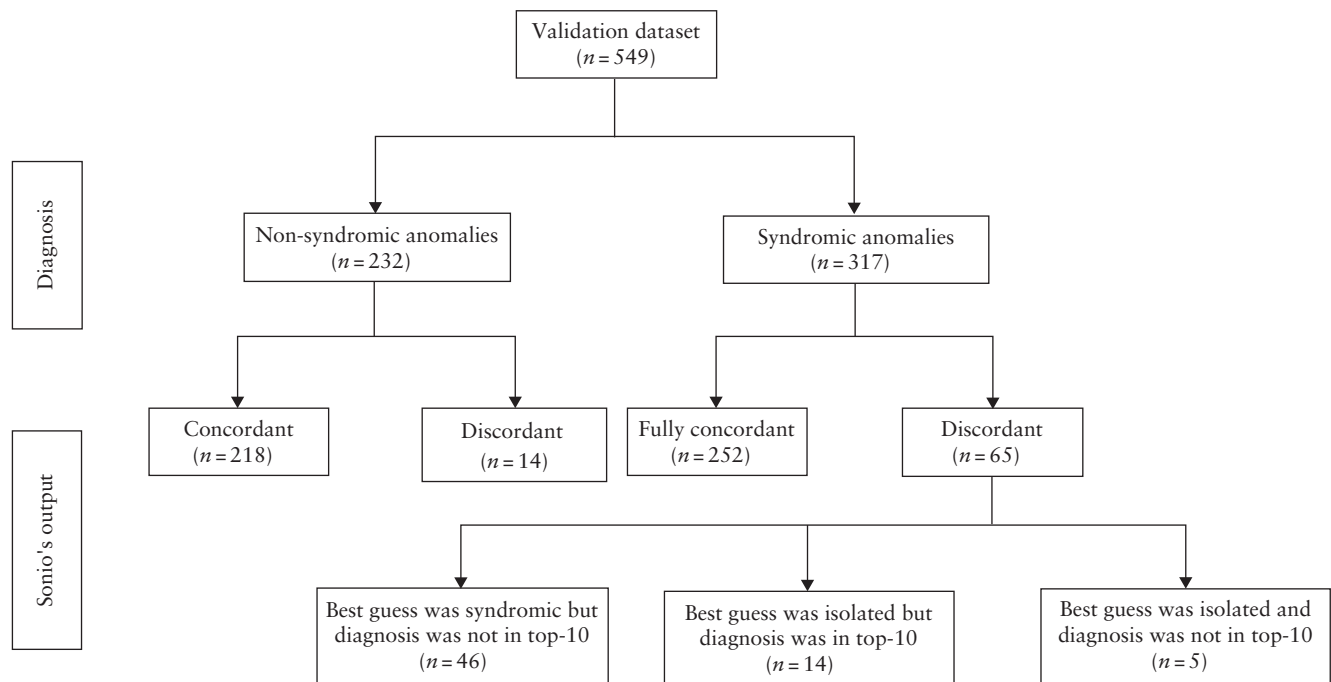

Figure 1 Flowchart summarizing validation dataset according to Sonio Expert's output. In syndromic cases, Sonio Expert's output was considered fully concordant when both best guess was a syndrome and ascertained disease was within top-10 list, and discordant otherwise. Fourteen non-syndromic discordant cases were considered false positives.

Table 2 Details of 14 cases of isolated anomaly labeled as syndromic by Sonio Expert (false positives)

Ascertained diagnosis	Input provided to Sonio Expert	Sonio Expert output	Qualitative interpretation of discordance
Pyelectasis	1. Pyelectasis 2. Hyperechogenic kidneys 3. Hydronephrosis 4. Unilateral obstruction of pyeloureteral junction	1. Turner syndrome 2. 22q11 deletion syndrome 3. Branchio-otorenal syndrome 4. Cat-eye syndrome 5. Wolf–Hirschhorn syndrome	Misunderstanding by software of causal dependencies of data provided
Pyelectasis	1. Hydroureter 2. Unilateral vesicoureteral reflux 3. Hydronephrosis 4. Pyelectasis	1. Megacystis–microcolon 2. Caudal duplication 3. Wolf–Hirschhorn syndrome 4. Rubinstein–Taybi syndrome 5. VACTERL/VATER association	Misunderstanding by software of causal dependencies of data provided
Pyelectasis	1. Hydroureter 2. Hydronephrosis 3. Pyelectasis	1. Megacystis–microcolon 2. Wolf–Hirschhorn syndrome 3. Rubinstein–Taybi syndrome 4. VACTERL/VATER association 5. Schinzel–Giedion syndrome	Misunderstanding by software of causal dependencies of data provided
Hydroureter	1. Hydroureter 2. Ascites	1. Mayer–Rokitansky syndrome 2. Non-syndromic association 3. Megacystis–microcolon 4. Cytomegalovirus 5. Schinzel–Giedion syndrome	Misunderstanding by software of causal dependencies of data provided
Alobar holoprosencephaly	1. Midline cleft lip 2. Small skull 3. Hypotelorism 4. Alobar holoprosencephaly	1. Holoprosencephaly–caudal dysgenesis syndrome 2. Trisomy 13 3. Holoprosencephaly–ectrodactyly–cleft lip/palate syndrome 4. Distal monosomy 13q 5. Smith–Lemli–Opitz syndrome	Misunderstanding by software of causal dependencies of data provided
Complete atrioventricular canal defect	1. Atrioventricular canal defect 2. Complete atrioventricular canal defect 3. Ventricular septal defect	1. Down syndrome 2. Non-syndromic association 3. Holt–Oram syndrome 4. CHARGE syndrome 5. Congenital rubella syndrome	Misunderstanding by software of causal dependencies of data provided
Pulmonary stenosis	1. Pulmonary stenosis 2. Ventricular septal defect	1. Noonan syndrome 2. Wolf–Hirschhorn syndrome 3. Turner syndrome 4. 8p inverted duplication/deletion 5. Cat-eye syndrome	Misunderstanding by software of causal dependencies of data provided
Postaxial foot polydactyly	1. Polydactyly postaxial 2. Postaxial foot polydactyly	1. Bardet–Biedl syndrome 2. Greig cephalopolysyndactyly 3. Joubert syndrome 4. Orofacio-digital syndrome 5. Trisomy 13	Misunderstanding by software of ontological dependencies of data provided
Unilateral cleft lip and palate	1. Unilateral cleft lip and alveolus 2. Unilateral cleft lip and palate	1. Diamond–Blackfan anemia 2. Kallmann de Morsier syndrome 3. Gorlin syndrome 4. Ectrodactyly–ectodermal dysplasia–cleft lip/palate syndrome 5. Fryns syndrome	Error when encoding clinical database. Two different and incompatible anomalies were provided to software
Ventriculomegaly	1. Ventriculomegaly 2. Mega cisterna magna 3. Absent septum pellucidum 4. Aplasia/hypoplasia of cerebellar vermis 5. Abnormal cerebral ventricle morphology	1. Non-syndromic association 2. Joubert syndrome 3. Coffin–Siris syndrome 4. Cytomegalovirus 5. PHACE syndrome	Misunderstanding by software of causal dependencies of data provided
Preterm prelabor rupture of membranes	1. Severe oligohydramnios 2. Short cervix	1. Cytomegalovirus 2. Triploidy 3. Non-syndromic association	Misunderstanding by software of causal dependencies of data provided
Two cases of semilobar holoprosencephaly, one case of lobar holoprosencephaly	Semilobar holoprosencephaly	Nothing, diagnosis screen displays 'Healthy'	Crash of algorithm

ware to identify a case as syndromic remained unaffected by the adjudication procedure: 94% (95% CI, 91–97%) and 96% (95% CI, 94–98%) for the full-phenotype and stepwise models, respectively (Table 3).

Reclassification of the 65 discordant syndromic cases by external expert adjudication is presented in Figure 2. Of these 65 discordant cases, the diagnosis was deemed unachievable based on the available phenotype in 44 (68%) cases, because the phenotype was either insufficient or atypical. In four cases, the adjudicators identified a synonym of the ascertained diagnosis in the top-10 list (3/46 cases for which the best guess was a syndrome and 1/5 for which the best guess was non-syndromic). Among the 51 syndromic cases for which the ascertained diagnosis was not in the top-10 list, adjudication concluded that the diagnosis could have been achieved based on the available phenotype in 15 (29%) cases (Table 4).

Given that 16 discordant cases for which the best guess was an isolated anomaly were deemed unachievable (Figure 2), the postadjudication sensitivity for the presence of a syndrome (irrespective of its nature) was 99% (95% CI, 98–100%) and 84% (95% CI, 82–86%) for the full-phenotype and stepwise models, respectively (Table 3). Excluding the 32 cases for which the diagnosis was not achievable from the group of 317 syndromic

cases, the postadjudication top-10 concordance was 93% (95% CI, 90–96%) using the full-phenotype model and 83% (95% CI, 81–86%) with the stepwise model. The postadjudication cumulative proportions of ascertained diagnoses within the top-10 list increased from 56% to 93% from the first to tenth rank with the full-phenotype model, and from 59% to 83% from the first to tenth

Table 4 Ascertained diagnosis in 15 syndromic cases for which incorrect classification by Sonio Expert was determined by expert adjudication to be a true error (i.e. diagnosis should have been made based on phenotype)

Disease	n
Trisomy 18	5
Triploidy	1
Fetal akinesia sequence	1
Fanconi anemia	1
L1 syndrome	1
DK phocomelia syndrome	1
Walker–Warburg syndrome	1
Achondroplasia	1
Atelosteogenesis Type 1	1
Cornelia de Lange syndrome	1
Osteogenesis imperfecta	1

Table 3 Sensitivity, specificity and top-10 concordance for Sonio Expert output, using 'full-phenotype' and 'stepwise' models, before and after adjudication

Emulation model	Sensitivity	Specificity	Top-10 list
Full phenotype			
Preadjudication	298/317 (94 (91–97))	218/232 (94 (91–97))	266/317 (84 (80–88))
Postadjudication	298/301 (99 (98–100))	218/232 (94 (91–97))	254/273 (93 (90–96))
Stepwise			
Preadjudication	815/991 (82 (80–85))	305/318 (96 (94–98))	806/991 (81 (79–84))
Postadjudication	815/972 (84 (82–86))	305/318 (96 (94–98))	781/939 (83 (81–86))

Data are given as *n/N* (% (95% CI)). Sensitivity and specificity refer to ability of software to differentiate syndromic from isolated case. Top-10 concordance describes ability of software to identify ascertained diagnosis in top-10 differential diagnoses list. For stepwise model, sample size is number of scenarios obtained by randomly changing first input symptom.
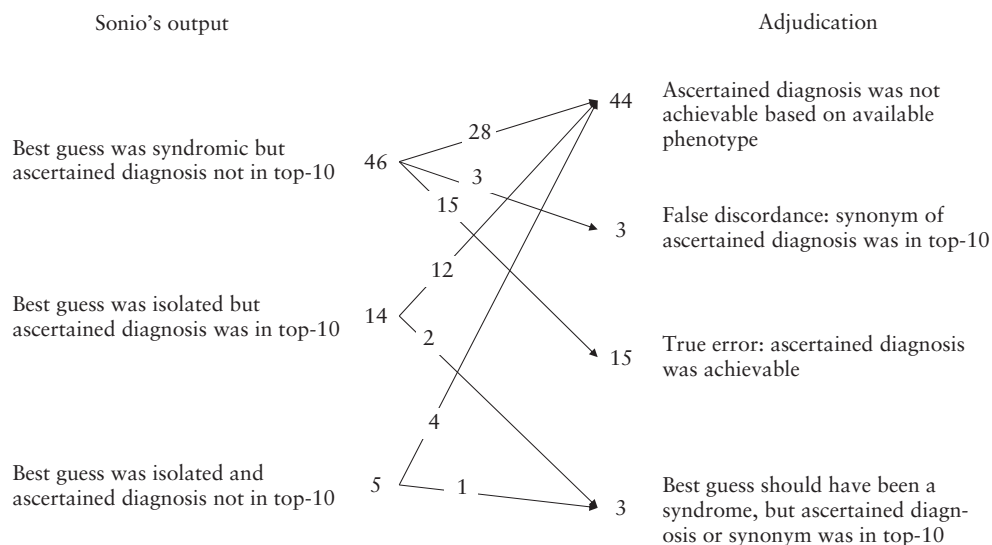

Figure 2 Schematic diagram showing reclassification of 65 discordant syndromic cases by expert adjudication.

rank using the stepwise model (Table 5). The stepwise approach required a mean of 13 queries to reach the final set of diagnoses.

DISCUSSION

Main findings

In this validation study using real-world data, the DSS appropriately identified a syndrome in > 95% of cases when the full phenotype was provided. Furthermore, the ascertained diagnosis was within the top-10 list in > 90% and > 80% of cases for the full-phenotype and stepwise approaches, respectively. These results show that Sonio Expert was remarkably robust to noise, given that, on average, half of the anomalies were atypical of the diagnosis. Both models also showed very high specificity (> 90%), demonstrating the ability of the software to identify causal relationships between a primary symptom and its consequences; the software appropriately identified a phenotype as non-syndromic even in the presence of multiple anomalies linked by causal relationships.

Interpretation

Given the growing complexity of medical knowledge and the increasing availability of data sources, DSSs have been developed to assist clinicians in their decision-making, particularly for the diagnosis of rare diseases[14,15]. However, to our knowledge, such systems have not been developed for the specific needs of fetal ultrasound. Online search engines that link given fetal ultrasound phenotypes to specific diseases exist[16], but such solutions implicitly assume that all possible fetal symptoms have been checked for and that the phenotype is therefore 'complete', which is usually not the case. Furthermore, such systems do not offer real-time assistance during an ultrasound examination. Similar engines exist for postnatal clinical phenotypes, such as 'Phenomizer' developed with HPO and Orphanet[17]. Faviez *et al.*[15] recently reviewed the performance of diagnostic DSSs for rare diseases. The systems presented in this review rely on *post-hoc* postnatal phenotyping and do not provide real-time assistance. In the nine systems for which performance was reported, the success rate of the top-10 diagnoses ranged from 32%[18] to 99%[19]. Sonio Expert, which incorporates the additional complexity of real-time processing, therefore emerges as an effective tool with a success rate in the highest performance range of the available DSSs for rare diseases. This is particularly relevant, given that reported performance rarely relies on real-world data for validation, using instead simulated/*in-silico* validation[20]. Furthermore, compared with postnatal phenotypes, the prenatal phenotype is often less informative for a number of reasons: diagnostic features may appear later in development or may not be present prenatally; neurological and neurodevelopmental features are obviously inaccessible *in utero*; and some symptoms, such as subtle

dysmorphic features, are beyond the capacity of prenatal ultrasound to detect.

The performance of Sonio Expert was understandably lower using a stepwise model, although it was successful for the detection of a syndromic association in over 80% of cases; this is possibly closer to what one could expect in a real-life clinical scenario than with the full-phenotype model. With the stepwise model, the complete phenotype may not have been recovered in the end, since only the symptoms that Sonio Expert asked to be checked are provided in the diagnostic pathway (Table 1). However, because each query is directed by the software's database, this model is less prone to run into atypical signs. This explains the slightly higher proportions of top-1 (i.e the best guess) and top-2 correct answers with the stepwise model compared with the full-phenotype model (Table 5).

Strengths and limitations

A key strength of this study is its use of a unique database from two large centers. Compared with real life, this database is enriched with more challenging non-chromosomal diagnoses, since chromosomal defects, generally picked up prenatally by CMA, do not require a postmortem examination. Secondly, two different aspects of the performance of Sonio Expert are described: the full-phenotype model describes the performance of the database and diagnostic search engine, while the stepwise model mimics the clinical use of the software. Finally, adjudication of discordant cases was performed by an independent panel of external experts.

Several limitations of this study should be acknowledged. Despite its size, the validation database did not cover all possible fetal conditions. It may also have been biased, owing to the preferential selection of a subset of conditions requiring postmortem examination. Moreover, although Sonio's database probably covers a large proportion of existing syndromes, it will never be exhaustive, given the rapidly growing number of diagnoses, despite efforts to keep it up-to-date. Finally, in

Table 5 Cumulative proportions (%) according to rank of ascertained diagnosis in top-10 list, for 'full-phenotype' and 'stepwise' models, before (Pre) and after (Post) adjudication, in cases with ascertained diagnosis of syndromic association ($n = 317$)

Rank	Full phenotype		Stepwise	
	Pre	Post	Pre	Post
1	47.9	55.7	55.6	58.7
2	59.0	67.4	66.1	68.4
3	67.2	74.7	70.5	72.7
4	71.3	79.5	72.8	75.0
5	74.1	82.4	74.7	76.7
6	77.6	86.1	76.3	78.3
7	79.8	88.3	78.4	80.2
8	82.3	91.2	79.2	80.9
9	82.6	91.6	80.3	82.1
10	83.9	93.0	81.3	83.2

real-life fetal ultrasound, the interpretation of the ultrasound image itself is up to the practitioner. Given that our study relied on postmortem and postnatal phenotyping, our setting does not incorporate potential human error in the interpretation of ultrasound imaging[21].

Clinical impact of this study

Knowledge of the precise prenatal phenotype of over 400 congenital disorders is beyond the capability of any individual physician/sonographer. Therefore, there is a gulf between the intractable complexity of rare diseases and the universal availability of powerful prenatal imaging technologies such as ultrasound. The medical uncertainty generated by this gulf, fueled by growing legal pressure, has led to an increasing number of ultrasound examinations per woman, without improving diagnostic performance[22] but increasing the cost of prenatal screening for fetal anomalies. Suspicion of a fetal anomaly is known to undermine the attachment of a woman to her pregnancy, with a significant and prolonged impact even when a favorable prognosis is reached[23].

The strengths of Sonio Expert lie in the quality of its constitutive database, the incorporation of contextual information and risk factors and its ability to work in real time at the patient's side, by prompting the practitioner to look for anomalies that could have been overlooked otherwise, while maintaining the diagnostic process within an acceptable number of steps. Through its use of real-world data, this study adds high external validity to this list of strengths.

Conclusions

This validation study suggests that Sonio Expert could improve perinatal care by efficiently providing complex and otherwise overlooked knowledge to care-providers involved in prenatal diagnosis. More specifically, the expected benefits of this software are: improving the diagnosis of rare fetal conditions; providing a standardized examination in case of an anomaly; limiting unnecessary repeat examinations and parental anxiety; and offering continuous medical education.

Disclosures

J.J.S., R.B., E.S., S.A., E.L.P. and Y.V. are cofounders of the SONIO company. S.R., F.L. and H.P.S.P. are employed by SONIO.

REFERENCES

1. EUROCAT [Internet]. [cited 2010 Sep 28]. Available from: http://www.eurocat-network.eu/.
2. Orphanet, The portal for rare diseases and orphan drugs [Internet]. [cited 2021 May 18]. Available from: http://www.orpha.net/consor/www/cgi-bin/index.php?lng=EN.
3. Guided screening and diagnosis of congenital malformations - sonio sonio [Internet]. [cited 2021 May 20]. Available from: https://www.sonio.ai/.
4. Adler-Milstein J, Chen JH, Dhaliwal G. Next-generation artificial intelligence for diagnosis: from predicting diagnostic labels to "wayfinding". *JAMA* 2021; **326**: 2467–2468.
5. Montani S, Striani M. Artificial intelligence in clinical decision support: a focused literature survey. *Yearb Med Inform* 2019; **28**: 120–127.
6. https://vimeo.com/794924328/bece255b4f.
7. CRAT – Centre de référence sur les agents tératogènes chez la femme enceinte [Internet]. [cited 2021 May 19]. Available from: https://www.lecrat.fr/.
8. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, Callahan TJ, Chute CG, Est JL, Galer PD, Ganesan S, Griese M, Haimel M, Pazmandi J, Hanauer M, Harris NL, Hartnett MJ, Hastreiter M, Hauck F, He Y, Jeske T, Kearney H, Kindle G, Klein C, Knoflach K, Krause R, Lagorce D, McMurry JA, Miller JA, Munoz-Torres MC, Peters RL, Rapp CK, Rath AM, Rind SA, Rosenberg AZ, Segal MM, Seidel MG, Smedley D, Talmy T, Thomas Y, Wiafe SA, Xian J, Yüksel Z, Helbig I, Mungall CJ, Haendel MA, Robinson PN. The human phenotype ontology in 2021. *Nucleic Acids Res* 2021; **49**: D1207–D1217.
9. Human Phenotype Ontology [Internet]. [cited 2021 May 19]. Available from: https://hpo.jax.org/app/.
10. Besson R. Decision making strategy for antenatal echographic screening of foetal abnormalities using statistical learning [Internet] [These de doctorat]. Université Paris-Saclay (ComUE); 2019 [cited 2021 May 19]. Available from: http://www.theses.fr/2019SACLX037.
11. Dhombres F, Jouannic J-M, Jaulent M-C, Charlet J. Choix méthodologiques pour la construction d'une ontologie de domaine en médecine prénatale. *21èmes Journées Francophones d'Ingénierie des Connaissances*, Jun 2010, Nîmes, France, 171–182.
12. Quinlan JR. Induction of decision trees. *Mach Learn* 1986; **1**: 81–106.
13. Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*. Taylor & Francis: New York, NY; 1984.
14. Gambhir S, Malik SK, Kumar Y. Role of soft computing approaches in healthcare domain: a mini review. *J Med Syst* 2016; **40**: 287.
15. Faviez C, Chen X, Garcelon N, Neuraz A, Knebelmann B, Salomon R, Lyonnet S, Saunier S, Burgun A. Diagnosis support systems for rare diseases: a scoping review. *Orphanet J Rare Dis* 2020; **15**: 94.
16. Porat S, de Rham M, Giamboni D, Van Mieghem T, Baud D. Phenotip - a web-based instrument to help diagnosing fetal syndromes antenatally. *Orphanet J Rare Dis* [Internet]. 2014 Dec 10 [cited 2021 May 14];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4268872/.
17. The Phenomizer (Orphanet) – Clinical Diagnostics with Similarity Searches in Ontologies [Internet]. [cited 2021 May 20]. Available from: http://compbio.charite.de/phenomizer_orphanet/.
18. Chen J, Xu H, Jegga A, Zhang K, White PS, Zhang G. Novel phenotype-disease matching tool for rare genetic diseases. *Genet Med* 2019; **21**: 339–346.
19. Zemojtel T, Köhler S, Mackenroth L, Jäger M, Hecht J, Krawitz P, Graul-Neumann L, Doelken S, Ehmke N, Spielmann M, Oien NC, Schweiger MR, Krüger U, Frommer G, Fischer B, Kornak U, Flöttmann R, Ardeshirdavani A, Moreau Y, Lewis SE, Haendel M, Smedley D, Horn D, Mundlos S, Robinson PN. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci Transl Med* 2014; **6**: 252ra123.
20. Reyna MA, Nsoesie EO, Clifford GD. Rethinking algorithm performance metrics for artificial intelligence in diagnostic medicine. *JAMA* 2022; **328**: 329–330.
21. Smith-Bindman R, Hosmer WD, Caponigro M, Cunningham G. The variability in the interpretation of prenatal diagnostic ultrasound. *Ultrasound Obstet Gynecol* 2001; **17**: 326–332.
22. Ferrier C, Dhombres F, Khoshnood B, Randrianaivo H, Perthus I, Guilbaut L, Durand-Zaleski I, Jouannic JM. Trends in resource use and effectiveness of ultrasound detection of fetal structural anomalies in France: a multiple registry-based study. *BMJ Open* 2019; **9**: e025482.
23. Api O, Demir HN, Api M, Tamer I, Orbay E, Unal O. Anxiety scores before and after genetic sonogram. *Arch Gynecol Obstet* 2009; **280**: 553–558.