Two-stage approach for risk estimation of fetal trisomy 21 and other aneuploidies using computational intelligence systems

Neocleous, A. C.; Syngelaki, A.; Nicolaides, K. H.; Schizas, C. N.

Published in:
29th World Congress on ultrasound in obstetrics and gynecology

DOI:
10.1002/uog.17558

29th World Congress on Ultrasound in Obstetrics and Gynecology

12-16 October 2019
Berlin, Germany

Ultrasound Obstet Gynecol 2018; **51**: 503–508
Published online 4 March 2018 in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.17558

Two-stage approach for risk estimation of fetal trisomy 21 and other aneuploidies using computational intelligence systems

A. C. NEOCLEOUS[1,2], A. SYNGELAKI[3], K. H. NICOLAIDES[3] and C. N. SCHIZAS[2]

[1]*Department of Intelligent Systems Group, Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands;* [2]*Department of Computer Science, University of Cyprus, Nicosia, Cyprus;* [3]*Harris Birthright Research Centre for Fetal Medicine, King's College Hospital, London, UK*

KEYWORDS: bioinformatics; cell-free DNA test; chromosomal abnormalities; computational intelligence; data normalization

ABSTRACT

Objective *To estimate the risk of fetal trisomy 21 (T21) and other chromosomal abnormalities (OCA) at 11–13 weeks' gestation using computational intelligence classification methods.*

Methods *As a first step, a training dataset consisting of 72 054 euploid pregnancies, 295 cases of T21 and 305 cases of OCA was used to train an artificial neural network. Then, a two-stage approach was used for stratification of risk and diagnosis of cases of aneuploidy in the blind set. In Stage 1, using four markers, pregnancies in the blind set were classified into no risk and risk. No-risk pregnancies were not examined further, whereas the risk pregnancies were forwarded to Stage 2 for further examination. In Stage 2, using seven markers, pregnancies were classified into three types of risk, namely no risk, moderate risk and high risk.*

Results *Of 36 328 unknown to the system pregnancies (blind set), 17 512 euploid, two T21 and 18 OCA were classified as no risk in Stage 1. The remaining 18 796 cases were forwarded to Stage 2, of which 7895 euploid, two T21 and two OCA cases were classified as no risk, 10 464 euploid, 83 T21 and 61 OCA as moderate risk and 187 euploid, 50 T21 and 52 OCA as high risk. The sensitivity and the specificity for T21 in Stage 2 were 97.1% and 99.5%, respectively, and the false-positive rate from Stage 1 to Stage 2 was reduced from 51.4% to ∼1%, assuming that the cell-free DNA test could identify all euploid and aneuploid cases.*

Conclusion *We propose a method for early diagnosis of chromosomal abnormalities that ensures that most T21 cases are classified as high risk at any stage. At the same time, the number of euploid cases subjected to invasive or cell-free DNA examinations was minimized through a routine procedure offered in two stages. Our method is minimally invasive and of relatively low cost, highly effective at T21 identification and it performs better than do other existing statistical methods. Copyright © 2017 ISUOG. Published by John Wiley & Sons Ltd.*

INTRODUCTION

First-trimester screening for trisomy 21 (T21) by a combination of maternal age, fetal nuchal translucency thickness (NT), serum free β-human chorionic gonadotropin (β-hCG) and pregnancy-associated plasma protein-A (PAPP-A) can detect about 90% of affected pregnancies at a false-positive rate (FPR) of 5%[1,2]. The performance of the first-trimester combined test can improve with an increase in detection rate (DR) to > 95% and a decrease in FPR to < 3% with the addition of the ultrasonographic markers of absent nasal bone, and abnormal flow in the ductus venosus and across the tricuspid valve[1,3–5].

A recent major improvement in performance of T21 screening has been achieved with analysis of cell-free DNA (cfDNA) in maternal blood, with a DR of > 99% and a FPR of < 0.1%[6]. However, universal screening by cfDNA testing as an alternative to the combined test would be expensive and ignore the other benefits of the combined test, such as early detection of many major fetal defects, diagnosis of multiple pregnancies and their chorionicity, and early prediction of pregnancy complications, such as pre-eclampsia, with the potential of prevention through prophylactic pharmacological intervention. The alternative to universal screening by

Correspondence to: Dr A. C. Neocleous, Department of Intelligent Systems Group, Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, 9712 CP Groningen, The Netherlands (e-mail: neocleous.andreas@gmail.com)

Accepted: 12 June 2017

cfDNA testing is a strategy of cfDNA analysis contingent on the results of first-line screening by the combined test. This approach retains the major advantages of cfDNA testing in increasing DR and decreasing FPR, but at a considerably lower cost than offering cfDNA testing to the whole population.

Artificial neural networks (ANNs), which are a specific simplified version of computational intelligence, are being applied increasingly in medicine and biological research[7–10]. Essentially, ANNs deal with mathematical algorithms implemented in software that learn from historical data and capture the knowledge and the internal dynamics that are contained in the data. Suitably trained models of computational intelligence approach the functionality of small biological neural clusters in a fundamental manner that mimics human-like behavior. They constitute the digitized model of the biological brain and can detect complex non-linear relationships between dependent as well as independent variables in a dataset, which are undetectable by the human brain. ANNs can learn from their input data, also known as training data. Such learning is achieved because ANN models can infer a function from observations and can subsequently use this function.

The objective of this study was to examine the potential value of ANN schemes in the stratification of risk for fetal T21 and other chromosomal abnormalities (OCA), incorporating the use of the combined test, additional first-trimester ultrasonographic markers and cfDNA testing, with the aim of achieving the highest possible DR at the minimum possible number of cfDNA and invasive tests.

METHODS

Study population

The study population was derived from women with singleton pregnancies attending the Fetal Medicine Unit at King's College Hospital, London (March 2006 to May 2015), the Obstetric Ultrasound Unit at University College London Hospital, London (April 2009 to July 2013) and the Fetal Medicine Unit at Medway Maritime Hospital, Gillingham (April 2010 to May 2015) for aneuploidy screening at 11 + 0 to 13 + 6 weeks' gestation. Maternal age and other demographic characteristics were recorded and a transabdominal ultrasound examination was performed for measurement of fetal crown–rump length and NT thickness and for assessment of the presence or absence of the fetal nasal bone, reversed A-wave in the ductus venosus and tricuspid regurgitation, by sonographers who had received the appropriate Fetal Medicine Foundation Certificates of Competence[1–4]. The pregnancy was dated according to the measurement of fetal crown–rump length[11]. Maternal blood was collected and automated machines that provide reproducible results within 30 min were used to measure serum PAPP-A and free β-hCG concentrations (Delfia Express System, Perkin Elmer, Waltham, MA, USA).

The best available method for establishing the presence of T21 and OCAs was prenatal fetal karyotyping by chorionic villus sampling or amniocentesis and postnatal karyotyping from neonatal blood. Absence of the target condition was established by either prenatal karyotyping or clinical examination of a phenotypically normal neonate.

ANN diagnostic system

A feed-forward network of neurons consisting of a number of layers that are connected to each other was built. The first layer is called the input layer and contains as many neurons as the input parameters. The last layer is called the output layer and contains one neuron. Other added layers, placed between the first and the last, are called hidden layers. A typical ANN architecture has one or two hidden layers. In every connection there is a weight and an activation function that represent the process in the synapses of cells in a biological brain. The weights are optimized in the training process by presenting all the examples several times and calculating the error, which is then used to adjust the weights through a learning algorithm. The number of repetitions is a parameter and is called epochs. The number of layers and neurons, the activation functions and the epochs are parameters that are predefined, in most cases empirically by the system designer.

An unknown case is evaluated by an ANN by presenting the parameters to the input layer. This information is passed through the layers by applying the appropriate weights and the transfer functions. The output value of the neural network in the last layer takes values in the range between 0 and 1 due to the sigmoid transfer function. To classify a case into positive or negative, a cut-off point is used that is applied to the output of the neural network. This was optimized to achieve the highest detection of T21 at the lowest FPR.

Data reduction for handling the class-imbalanced effect

The vast majority of cases in our dataset were euploid, creating a high imbalance between the two classes, normal and abnormal. This is known as the class-imbalanced problem[12]. More precisely, from the total number of 72 654 cases used in the training set, 72 054 were euploid and 600 were aneuploid. In medical data, this is a typical finding, as abnormalities are less common, thus resulting in an imbalance between the normal and abnormal classes[13–15].

In machine learning, most of the supervised techniques for classification are not able to generalize the data, and the performance of the classifiers is low when the training set consists of imbalanced data. For instance, the Bayesian classifier uses the population of each class to estimate the posterior probabilities. In the case of class-imbalanced populations, the probability of an unknown case will be biased towards the majority class. Similarly, the popular classifier of support vector machines performs poorly with

imbalanced populations[16]. ANNs adjust their weights based on a classification error, as explained above. As the error is calculated globally for both normal and abnormal cases, a false-negative classification has equal impact as a false-positive classification. In practice, however, a false-negative classification has higher importance due to the small population. In other words, in our dataset, a false-positive classification had a percentage of $1/72\,054 = \sim 0\%$, whereas a false-negative classification had a percentage of $1/600 = 0.1\%$.

In the literature, a lot of work has been carried out on generating a balanced set from an imbalanced dataset for classification purposes[17–19]. This is typically carried out by either oversampling the minority class[20] or by downsampling the majority class[21]. For the problem under study, we applied both approaches and found that downsampling the majority class yielded better results.

Thus, first we generated a cluster map of the entire euploid population, using the k-means[22] algorithm with five prototypes, assuming that the normal cases have their own subclusters. Then, we computed the prototype vector for each subcluster using the k-means algorithm and selected representative cases around this vector. In this way, the result was a reduced training set for the euploid class that represents the entire euploid population. A more detailed explanation of this approach can be found in our previous work[23].

Cross-validation

A standard procedure was followed for testing the performance of our system. Typically, in machine learning, three different datasets are used for building and verifying a system, namely the training set, the validation set and the test set. The training set consists of known cases that are used for the learning procedure of the ANN and the validation set consists of known cases that are used to assess the learning performance of the system during the training phase. The test set is usually another dataset with labels known only to the doctor. When the test set is given for testing without knowing the state of each case, it is called a blind set.

The training and test sets are randomly chosen from a dataset with a percentage of 70% and 30% for training and validation sets, respectively, and the experiments are repeated three times with different sets. This procedure is called three-fold cross-validation. Similarly, the 10-fold cross-validation is typically carried out by using 90% of the dataset for training and 10% for validation. Another popular procedure is the leave-one-out cross-validation in which the training set consists of the total population and only one case is used for validation. This procedure is repeated as many times as the size of the dataset and the results are often presented with statistical terms, such as average and standard deviation. This is carried out to ensure that the results are consistent, even though the training sets are different.

The training set used in this study consisted of euploid cases and cases with T21, trisomy 18, trisomy 13, triploidy, Turner and OCA. A three-fold cross-validation with random selection was applied, and in each run, a 100% true-positive rate was achieved for T21 at a FPR of $< 5\%$, for both training and validation sets. After the development of the appropriate ANNs, another blind set was used to evaluate the performance of our system and produce the results presented in this paper.

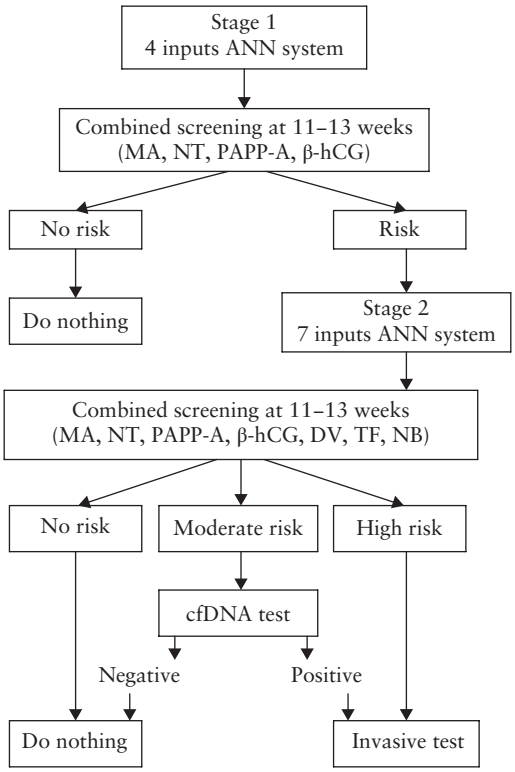

Figure 1 Overview of proposed two-stage approach for stratification of risk and diagnosis of fetal trisomy 21 and other chromosomal abnormalities. In Stage 1, each case undergoes prenatal examination for estimation of risk for fetal aneuploidy and is classified as risk or no risk. Risk cases are reassessed in Stage 2 and classified subsequently as no risk with no further action required, moderate risk with recommendation for cell-free DNA (cfDNA) testing or high risk with recommendation for invasive testing, in order to reach a diagnosis. β-hCG, β-human chorionic gonadotropin; ANN, artificial neural networks; DV, ductus venosus; MA, maternal age; NB, nasal bone; NT, nuchal translucency; PAPP-A, pregnancy-associated plasma protein-A; TF, tricuspid flow.

Stratification of risk

In this study, a two-stage approach was used for stratification of risk and the diagnosis of fetal T21 and OCA (Figure 1).

In Stage 1, four neurons were used in the input layer, representing the maternal age in years, serum free β-hCG, PAPP-A and NT in mm. The output was binary: no risk and risk for aneuploidy. The no-risk group from Stage 1 should not require any further testing and the group at risk for aneuploidy was subjected to Stage 2 screening.

In Stage 2, seven neurons were used in the input layer, representing the maternal age, free β-hCG, PAPP-A, NT, nasal bone (present or absent), ductus venosus flow

Table 1 Composition of whole population and training and blind sets used in study

Dataset	*Euploid*	*Trisomy 21*	*Trisomy 18*	*Trisomy 13*	*Triploidy*	*Turner*	*Other aneuploidy*
Total	108 112	432	166	56	35	63	118
Training	72 054 (reduced to 5002*)	295	115	37	29	45	79
Blind	36 058	137	51	19	6	18	39

Data are given as *n*. *Reduced by downsampling to balance training set[21].

(positive or negative A-wave) and tricuspid flow (present or absent tricuspid regurgitation). The output was no risk, moderate risk and high risk for aneuploidy. The no-risk group should not require any further testing, the moderate-risk group should have only cfDNA testing and the high-risk group would be strongly advised to undergo invasive testing without going through cfDNA testing.

RESULTS

Study population

The composition of the whole population of the training and validation sets used in the study is shown in Table 1. In total there were 108 112 euploid and 870 aneuploid pregnancies, including 432 cases of T21, 166 of trisomy 18, 56 of trisomy 13, 35 of triploidy, 63 of Turner syndrome and 118 of other aneuploidies. The training set contained 72 054 euploid pregnancies that were reduced to 5002, as explained in the Methods section, 295 cases of T21 and 305 of other aneuploidies, whereas the validation set contained 36 058 euploid pregnancies, 137 cases of T21 and 133 of other aneuploidies.

Stratification of risk

In Stage 1 we used four markers (maternal age, PAPP-A, β-hCG and NT) as inputs to the ANN. From the 36 328 pregnancies in the blind set, 17 532 (48.3%) were classified as no risk and 18 796 (51.7%) were allocated to the risk group that was subsequently assessed in Stage 2 (Figure 2).

Three additional markers were included in Stage 2, namely the ductus venosus, tricuspid flow and nasal bone. Of the 137 T21 cases in the blind set, 50 cases were allocated to the high-risk group, 83 to the moderate-risk group and two to the no-risk group. The sensitivity and the specificity for T21 in Stage 2 were 97.1% and 99.5%, respectively, assuming that the cfDNA test can identify all the euploid and aneuploid cases. Furthermore, 187 of the euploid pregnancies were allocated to the high-risk group, 10 464 to the moderate-risk group and 7895 to the no-risk group (Figure 2). The FPR from Stage 1 to Stage 2 was reduced from 51.4% to ~1%, assuming that the cfDNA test can identify all the euploid and aneuploid cases.

In addition to the diagnosis of T21, our method achieved high accuracy in detecting OCA. The validation set contained 133 pregnancies with aneuploidies other than T21; 18 of these were allocated to the no-risk group in Stage 1. Of the 115 cases allocated to the risk group in Stage 1, two, 61 and 52 were allocated, respectively, to the no-risk, moderate-risk and high-risk groups in Stage 2.

DISCUSSION

The findings of this study demonstrate the potential value of ANN schemes in the prediction of T21 and other aneuploidies from ultrasonographic and biochemical markers at 11–13 weeks' gestation. We used multilayer feed-forward neural systems because these are considered to be the most suitable from the point of view of satisfactory generalization and diagnostic yield[24]. Essentially, a multilayer network of neurons was built and adjusted according to a set of parameters for each case of either aneuploidy or euploidy in order to maximize the correct identification of each group.

ANNs have the ability to handle non-linear structures by using multiple hidden layers. Furthermore, assumptions about statistical concepts such as distributions, mean and standard deviation values are not needed. In addition to the above advantages, they can learn to recognize patterns in data and they have been used widely for medical tasks, such as image recognition, for several diseases.

Here we present a two-stage approach for the estimation of the risk of aneuploidy. In both stages, it was ensured, by adjustment of the cut-off point accordingly, that only a minimum number of T21 cases were classified as euploid (i.e. 97% sensitivity). At the same time, the studies were focused to minimize the FPR to the lowest possible. We have validated our results using a test set of a total number of 36 058 euploid, 137 T21 and 133 OCA cases. In Stage 2, 10 608 pregnancies were allocated to the moderate-risk group and consequently 29% of the total population of 36 328 pregnancies would require cfDNA testing. However, the proportion of euploid cases that would be advised to undergo invasive testing was less than 1% of the total population. The values here assume that cfDNA detects accurately all euploid and aneuploid cases. Although cfDNA testing performance for T21 approaches this assumption, for OCAs, this is not accurate.

We report higher classification results than the state-of-the-art statistical mixture model that is currently used as a classifier. For making an accurate comparison between our method and the standard first-trimester serum plus ultrasound screening test, we compared the 95% DR for T21 that is reported in the literature at 5% FPR. Adjusting a cut-off point at the value of 0.45, as explained in the Methods section, DRs of 94.2% and 79.5% were achieved for T21 and OCA, respectively, at

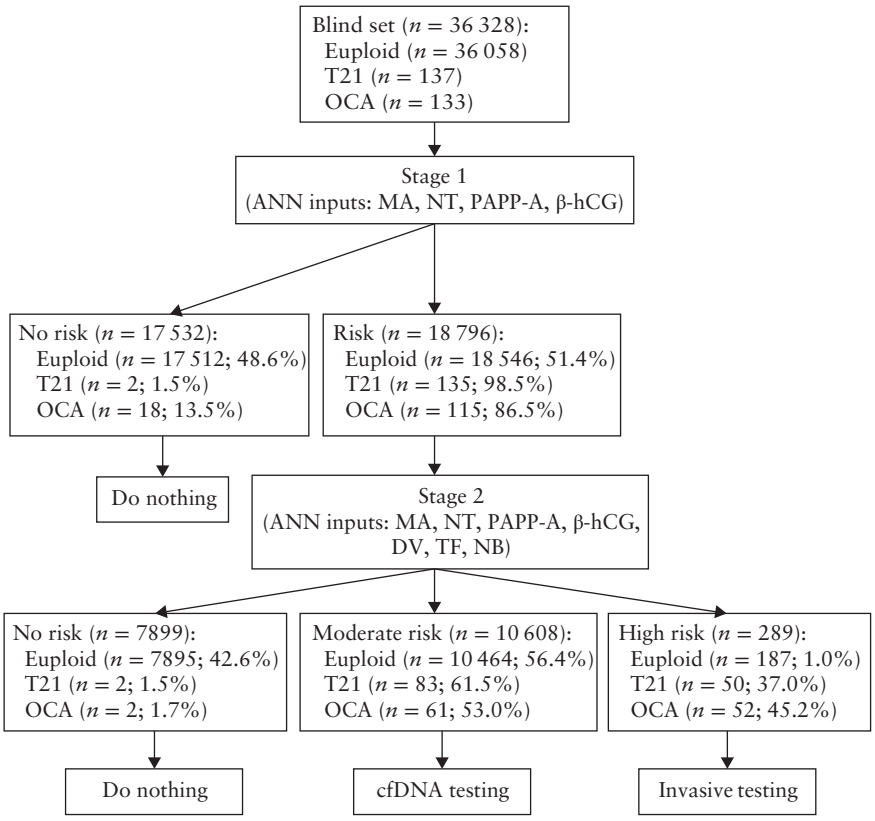

Figure 2 Classification of cases in blind set according to risk for chromosomal abnormality by artificial neural network (ANN) in Stages 1 and 2. Percentages are calculated against respective number of cases in previous stage. β-hCG, β-human chorionic gonadotropin; cfDNA, cell-free DNA; DV, ductus venosus; MA, maternal age; NB, nasal bone; NT, nuchal translucency; PAPP-A, pregnancy-associated plasma protein-A; OCA, other chromosomal abnormality; T21, trisomy 21; TF, tricuspid flow.

a FPR of 1.2%. Therefore, for the same DR of 95% for T21, we achieved a significantly lower FPR.

Our proposed methodology has the potential to be used in a real-time application in medical centers, as it returns immediate results during a regular visit of the pregnant woman, thus reducing the time and cost for additional examinations. Moreover, it can have a built-in learning mechanism, which will add continually to the knowledge acumen of the system while new identified cases are added into the system. Currently, updates are carried out manually at certain intervals. As the application is installed in a normal computer, the doctor could easily use it and the validation of the cases will be carried out at no cost.

One drawback of our method is that it does not classify correctly, at both stages, the euploid cases as no risk. About 51% of the euploid population after Stage 1 will be requested to access sonographers with training in the assessment of certain markers that are not routinely assessed by everyone. Access to such doctors is relatively limited for the whole advantaged population. Furthermore, in a proportion of patients, some of those markers cannot be obtained successfully even by experienced doctors, due to fetal position, for example. About 30% of the euploid population in Stage 2 is classified as moderate risk, thus resulting in many cases requiring further testing, such as cfDNA testing. This fact makes the proposed method weak in terms of cost-effectiveness, but it is better than the entire euploid population undergoing cfDNA testing.

Certainly, more work needs to be done to improve further the DR of the OCA cases. Our outcome assessments in Stage 2 are based on prenatal karyotypes of fetuses that screened positive and postnatal karyotypes of fetuses that are not phenotypically normal. Several OCAs will not be picked up by that type of assessment. The cfDNA testing does not target, and thus detect, several OCAs, while some of them are not diagnosable by phenotype assessment at birth. Furthermore, none of the low-risk and moderate-risk cases was considered phenotypically normal at birth. Several of the cases considered here as euploid would actually be carriers of OCAs not targeted by cfDNA and/or with an apparently normal phenotype at birth. In Tables S1–S3, each of the OCAs that were allocated to the no-risk group in Stage 1 and to the no-risk and moderate-risk groups in Stage 2 is described. The first ones would only get a karyotype if they had an abnormal phenotype at birth and the second ones would have only cfDNA testing, which does not target all OCAs.

In this study we ensured that only 3% of T21 cases would be born unexpectedly with this aneuploidy (DR 97%). We were able to limit the number of cfDNA tests in comparison with a clinical approach offering this test as a first tier to every woman, while keeping the

number of invasive tests as low as possible with such an approach.

Research work is now focusing on building models that associate the risk for aneuploidy with pre-eclampsia and other pregnancy complications. Preliminary results are already available.

ACKNOWLEDGMENTS

We thank Dr C. Neocleous and Dr K. Neokleous from the Computational Intelligence Laboratory of the University of Cyprus for their help and support and Prof. N. Petkov from the Department of Intelligent Systems Group, Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands for reviewing the paper and his valuable guidance.

REFERENCES

1. Nicolaides KH. Screening for fetal aneuploidies at 11 to 13 weeks. *Prenat Diagn* 2011; **31**: 7–15.
2. Santorum M, Wright D, Syngelaki A, Karagioti N, Nicolaides KH. Accuracy of first trimester combined test in screening for trisomies 21, 18 and 13. *Ultrasound Obstet Gynecol* 2017; **49**: 714–720.
3. Kagan KO, Cicero S, Staboulidou I, Wright D, Nicolaides KH. Fetal nasal bone in screening for trisomies 21, 18 and 13 and Turner syndrome at 11–13 weeks of gestation. *Ultrasound Obstet Gynecol* 2009; **33**: 259–264.
4. Maiz N, Valencia C, Kagan KO, Wright D, Nicolaides KH. Ductus venosus Doppler in screening for trisomies 21, 18 and 13 and Turner syndrome at 11–13 weeks of gestation. *Ultrasound Obstet Gynecol* 2009; **33**: 512–517.
5. Kagan KO, Valencia C, Livanos P, Wright D, Nicolaides KH. Tricuspid regurgitation in screening for trisomies 21, 18 and 13 and Turner syndrome at 11 + 0 to 13 + 6 weeks of gestation. *Ultrasound Obstet Gynecol* 2009; **33**: 18–22.
6. Gil MM, Accurti V, Santacruz B, Plana MN, Nicolaides KH. Analysis of cell-free DNA in maternal blood in screening for aneuploidies: updated meta-analysis. *Ultrasound Obstet Gynecol* 2017; **50**: 302–314.
7. Patel JL, Goyal RK. Applications of artificial neural networks in medical science. *Curr Clin Pharmacol* 2007; **2**: 217–226.
8. Schnorrenberg F, Tsapatsoulis N, Pattichis CS, Schizas CN, Kollias S, Vassiliou M, Adamou A, Kyriacou K. A modular neural network system for the analysis of nuclei in histopathological sections. *IEEE Eng Med Biol Mag* 2000; **19**: 48–63.
9. Schnorrenberg F, Pattichis CS, Kyriacou K, Schizas CN. Computer-aided detection of breast cancer nuclei. *IEEE Trans Inf Technol Biomed* 1997; **1**: 128–140.
10. Schizas CN, Pattichis CS. Learning systems in biosignal analysis. *BioSystems* 1997; **41**: 105–125.
11. Robinson HP, Fleming JE. A critical evaluation of sonar crown rump length measurements. *Br J Obstet Gynaecol* 1975; **182**: 702–710.
12. Japkowicz N, Stephen S. The class imbalance problem: a systematic study. *Intell Data Anal* 2002; **6**: 429–449.
13. Klement W, Wilk S, Michalowski W, Farion KJ, Osmond MH, Verter V. Predicting the need for CT imaging in children with minor head injury using an ensemble of Naive Bayes classifiers. *Artif Intell Med* 2012; **54**: 163–170.
14. Boughorbel S, Al-Ali R, Elkum N. Model comparison for breast cancer prognosis based on clinical data. *PloS One* 2016; **11**: e0146413.
15. Sheikhi G, Altınçay H. The cost of type II diabetes mellitus: a machine learning perspective. *XIV Mediterranean Conference on Medical and Biological Engineering and Computing* 2016. Springer: Berlin, 2016; 818–821.
16. Datta S, Das S. Near-Bayesian support vector machines for imbalanced data classification with equal or unequal misclassification costs. *Neural Networks* 2015; **70**: 39–52.
17. Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA Tourassi GD. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. *Neural Networks* 2008; **21**: 427–436.
18. Wilk S, Stefanowski J, Wojciechowski S, Farion KJ, Michalowski W. Application of preprocessing methods to imbalanced clinical data: an experimental study. *Information Technologies in Medicine*. Springer: Berlin, 2016; 503–515.
19. He H, Ma Y (eds). *Imbalanced Learning: Foundations, Algorithms, and Applications*. John Wiley & Sons: New York, 2013.
20. Pérez-Ortiz M, Gutiérrez PA, Tino P, Hervás-Martínez C. Oversampling the minority class in the feature space. *IEEE Trans Neur Net Learn Syst* 2015; **27**, 1947–1961.
21. Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. *ICML* 1997; **97**: 179–186.
22. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proc Fifth Berkeley Symp Math Stat Prob* 1967; **1**: 281–287.
23. Neocleous A, Nicolaides K, Schizas C. Intelligent noninvasive diagnosis of aneuploidy: raw values and highly imbalanced dataset. *IEEE J Biomed Health Informatics* 2017; **21**: 1271–1279.
24. Neocleous AC, Nicolaides KH, Schizas CN. First trimester noninvasive prenatal diagnosis: a computational intelligence approach. *IEEE J Biomed Health Informatics* 2015; **20**: 1427–1438.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

Table S1 Other chromosomal abnormalities classified as no risk in Stage 1

Table S2 Other chromosomal abnormalities classified as low risk in Stage 2

Table S3 Other chromosomal abnormalities classified as moderate risk in Stage 2

