Intelligent Noninvasive Diagnosis of Aneuploidy: Raw Values and Highly Imbalanced Dataset

Andreas C. Neocleous, Kypros H. Nicolaides, and Christos N. Schizas, *Senior Member, IEEE*

***Abstract*—The objective of this paper is to introduce a noninvasive diagnosis procedure for aneuploidy and to minimize the social and financial cost of prenatal diagnosis tests that are performed for fetal aneuploidies in an early stage of pregnancy. We propose a method by using artificial neural networks trained with data from singleton pregnancy cases, while undergoing first trimester screening. Three different datasets[1] with a total of 122 362 euploid and 967 aneuploid cases were used in this study. The data for each case contained markers collected from the mother and the fetus. This study, unlike previous studies published by the authors for a similar problem differs in three basic principles: 1) the training of the artificial neural networks is done by using the markers' values in their raw form (unprocessed), 2) a balanced training dataset is created and used by selecting only a representative number of euploids for the training phase, and 3) emphasis is given to the financials and suggest hierarchy and necessity of the available tests. The proposed artificial neural networks models were optimized in the sense of reaching a minimum false positive rate and at the same time securing a 100% detection rate for Trisomy 21. These systems correctly identify other aneuploidies (Trisomies 13&18, Turner, and Triploid syndromes) at a detection rate greater than 80%. In conclusion, we demonstrate that artificial neural network systems can contribute in providing noninvasive, effective early screening for fetal aneuploidies with results that compare favorably to other existing methods.**

***Index Terms*—Bioinformatics, chromosomal abnormalities, computational intelligence, data normalization, fetal aneuploidies, imbalanced data, machine learning, noninvasive prenatal diagnosis.**

I. Introduction

THE early diagnosis of fetal aneuploidies in the first trimester of pregnancy can be achieved with amniocentesis or chorionic villus sampling (CVS). However, such methods are invasive and they carry a risk of infections, fetal damage during the examination and miscarriage rates of about 0.4% for amniocentesis and 1.1% for CVS [1].

In Europe, the cost of an amniocentesis test varies between 300 and 1 000 euros. The cost of the loss of a life due to amniocentesis is tremendous. Therefore, it is important to reduce the false positive rate (FPR) as much as possible, but at the same to be able to detect all, or most of the aneuploidies. Previously, the overall cost for the detection was much higher because pregnant women were advised to go for amniocentesis based only on their age, e.g., greater than 35 years old, which not only increased the FPR beyond 25% and the risk of unexpected aneuploidy births much greater than zero; since in many cases, the cost for the invasive test could not be afforded and/or the risk for miscarriage could not be taken by the parents.

As alternative, noninvasive methods have been proposed by Snijders *et al.* [2], [3], Kagan *et al.* [4], [5], and Spencer *et al.* [6], [7]. Essentially, a risk for aneuploidy is estimated based on a prenatal examination test that is performed to every pregnant woman in the first trimester of pregnancy. In the literature, the most relevant markers from the prenatal examination are the following: maternal age, serum free β-hCG, pregnancy-associated plasma protein-A, nuchal translucency (NT) thickness, nasal bone, tricuspid flow, and ductus venosus flow (DV). A statistical mixture model (SMM) is used in [2]–[10] as an estimator for the risk of Trisomy 21 (T21). Outcomes that are rated as "high risk" are suggested to follow invasive test. Currently, the detection rate (DR) for the T21 of the abovementioned methods is 95% at a 5% FPR.

In the last years, another noninvasive method has gained particular attention in the scientific community. A sample from the maternal blood is used to isolate the plasma via double centrifugation. Then, the circulating cell-free DNA is sequenced from the maternal plasma, using the state-of-the-art equipment. The density of the DNA sequences are normalized and distributed for every chromosome separately for euploid, T21, and T18. A standard statistical classification technique such as Z-transform or t-test is applied to the euploidy and aneuploidy distributions of the chromosomes 21 and 18 to estimate the probability for aneuploidy of an unknown case [11]–[15].

In this paper, we propose the use of a machine learning approach by using artificial neural networks (ANNs) that have the ability to identify patterns from a training dataset, in a similar manner to the biological function of the human brain [16]. In our previous work [17], we have showed that ANNs can achieve 100% DR for T21 at 3.6% FPR.

Manuscript received April 22, 2016; revised July 23, 2016 and August 18, 2016; accepted September 6, 2016. Date of publication December 22, 2016; date of current version September 1, 2017.

A. C. Neocleous is with the Department of Intelligent Systems Group, Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen 9712 CP, The Netherlands, and also with the Department of Computer Science, University of Cyprus, Nicosia 2109, Cyprus (e-mail: neocleous.andreas@gmail.com).

K. H. Nicolaides is with the Harris Birthright Research Center for Fetal Medicine, King's College Hospital, London SE5 9RS, U.K. (e-mail: Kypros@fetalmedicine.com).

C. N. Schizas is with the Department of Computer Science, University of Cyprus, Nicosia 2109, Cyprus (e-mail: schizas@ucy.ac.cy).

Digital Object Identifier 10.1109/JBHI.2016.2608859

[1]The datasets can become available for academic purposes by communicating directly with the authors.

A. Statistical Mixture Model

A prenatal examination is performed to the pregnant woman between the 9th + 3 and 11th week + 6 days of gestation. From the maternal blood, the concentrations of the two biochemical markers: 1) pregnancy-associated plasma protein-A (PAPP-A); and 2) the serum free β-hCG are measured. Fetal ultrasonographic markers include the levels of appearance of a subcutaneous collection of fluid behind the fetal neck, called NT. The amount of the NT is statistically increased in fetuses with T21 [4]–[6]. Other sonographic markers are measured such as the crown rump length (CRL) and the nasal bone (NB). The CRL is a physical measurement of the length in millimeters between the neck and the bottom of the buttocks of the fetus. The NB indicates the presence or the absence of the fetal NB and the obstetrician marks it at the time of the examination as normal or abnormal, respectively.

It has been proposed in [4]–[6], [8]–[10] that the biochemical markers β-hCG and PAPP-A and the NT increase the separability strength between euploid and aneuploid. Additionally, the values of the biochemical markers are normalized with their multiples of their medians (MoM). The MoMs is a data normalization method that has been found effective in medical data [4], [5] and it is a measure of how many times an individual test result deviates from the median. The term "multiples" refers to this measure. The MoMs are calculated as follow: First, the data are clustered in three categories based on the gestational age at the time of the examination:

1) 9–10;
2) 10–11; and
3) 11–12 weeks of gestation.

Then, for every category the median value of the biochemical markers for the euploid cases is calculated. Finally, the raw observation is divided by the respective median value of the specific gestation age. The NT did not respond positively to the MoM values and it is transformed into the delta NT that is a measure of the deviation from its euploid zero median.

The classification is done as follow: For every marker, a risk is estimated based on the Gaussian distribution of the respective marker values and it is multiplied with the maternal age related risk to yield a final result. The DV and the tricuspid valve flow (TF) may be used to increase the DR and reduce the FPR [9], [10].

B. Cell-Free Fetal DNA Test

In the literature, the most promising results within the noninvasive methods are achieved with the cell-free fetal DNA test. In [11], it is reported that a perfect isolation of 300 euploid and 50 cases of T21 is achieved, together with a 98% DR for the trisomy 18 (T18). This method returned no result in three euploid cases (about 1% of the population). Another recent study [12] reports 100% DR of 14 cases of T21 at 0% FPR with 26 euploid. The results in [13] show that significantly lower FPR can be achieved compared to the SMM method that is currently used. Particularly 146 958 cases have been studied including 726 T21, 170 T18, and 22 trisomy 13 (T13) for which outcome data were available in 112 669 (76%). The overall sensitivity and specificity are 99.02% and 99.86%, respectively. Other studies [14] and [15] report similar results.

Even though the results of the *cell-free fetal DNA* approach are significantly better than other noninvasive methods, we identify some drawbacks. First, the test does not yield immediate results. It requires specialized doctors, expensive equipment, and laboratories. Therefore, it is a method that cannot be applied in one visit. Another drawback of this method is that the DNA sequence is statistically visualized and the classification is done by using a threshold on the probability of aneuploidy risk. This means that the results need to be cross validated in several standard ways, e.g., three-fold or leave-one-out cross validation to make sure that are consistent. Finally, the other chromosomal abnormalities (OCA) such as Turner syndrome or triploidy are not identified. However, a combination between the existing noninvasive methods may yield to an optimized solution to the problem. A combination of two noninvasive methods, the prenatal screening, and the maternal blood cell-free fetal DNA testing is proposed in [15].

C. Machine Learning

ANNs have been widely used in medical applications for the prediction of cancer [18]–[20], Parkinson disease [21], and other serious diseases [22]–[24]. The major difference between statistical methods and ANNs in classification is that ANNs have the ability to learn and store information by examples that are presented one by one. In other words, statistical information such as the distribution, the mean, and standard deviation values are not significant. This is done in a similar process as the biological neural networks process information in the brain.

The research question of this study is to examine the potential value of ANNs in the prediction of the risk for T21 and OCA from ultrasonographic and biochemical markers at the early stage of gestation. Additionally, we make experiments to test the possible contribution of normalized data over the raw data. Furthermore, we question if the use of balanced training sets between euploid and aneuploid achieve more reliable and consistent results. Finally, we explore the contribution of different combinations of input markers. The objective of this study is to build a system that will identify 100% of the T21 at the lowest FPR possible, using the most reliable type of markers for the training of the ANNs, being raw or normalized, balanced or imbalanced training sets, and the combination of the input markers.

In Section II, we present our methodology that includes a description and analysis of the available data, the ANN structures, the cross-validation approach for testing the models, the normalization of the data, and the procedure for creating balanced training sets. In Section III, we present our results and we conclude with Sections IV–VI.

II. Methods

Three different datasets are provided by the Fetal Medicine Foundation (FMF) and used in this study. The first (*Dataset A*), consists of 51 001 cases of pregnant women that followed a prenatal examination within the first trimester of pregnancy, and

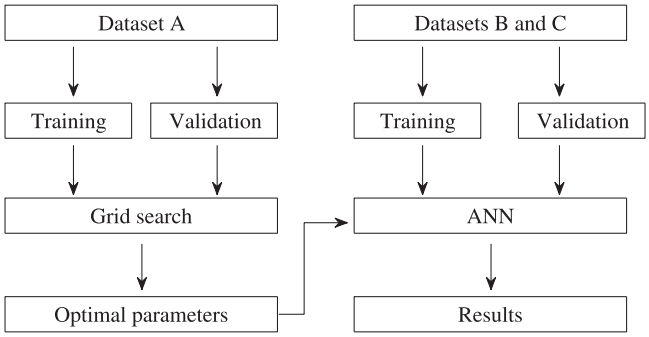

Fig. 1. All of the three datasets are divided into training and validation sets. The training set of the *Dataset A* is used to perform a grid search over different parameters of the ANN. We use the parameters of the network that yielded the best results on the validation set of *Dataset A* to perform experiments using the *Datasets B* and *C*.

similarly the second (*Dataset B*) and the third (*Dataset C*) of 29 999 and 42 329 cases, respectively. All the samples were very carefully collected and thoroughly tested, and they are maybe the largest complete dataset in existence today for this kind of study.

We used part of *Dataset A* as training set for performing a grid search of ANNs over a set of parameters, such as the hidden units, activation functions, and training epochs, as shown in Fig. 1. Then, we selected the ANN that yielded the best results on the remaining cases that formed the test set of *Dataset A*. All of our experiments were then done by using *Datasets B* and *C* that we call "testing datasets." The ANNs were built by using the parameters of the ANN that was identified as the optimal in the grid search.

We proceeded with our research approach by implementing three different experiments. The first experiment was done for identifying the optimal combination of markers that are needed as input to our system. We created eight groups of markers and we built an ANN for each one of them to compare their performance. The second experiment was done for testing whether the use of normalized data values for the markers outperform the use of raw values. The third experiment dealt with a more technical question concerned with the imbalanced nature of the datasets, which is due to the very low percentage of aneuploidies in the datasets. A data reduction technique was carried out for reducing the population of euploid cases during the training phase. Several representative ANNs were developed and tested, both with balanced and imbalanced datasets and the results were compared.

A. Data

The populations of euploid and aneuploid of each dataset are shown in Table I. The vast majority of the cases are euploid creating a highly imbalanced situation between the euploid and aneuploid cases.

The number of markers/features that were available for our dataset is 22. Most of them are related to the physiological and historical data of the pregnant woman, such as the history of aneuploidy in previous pregnancies, smoking or drug habits, symptoms of hypertension, way of conception, ethnicity, etc. Other markers of apparently greater importance since they are taken during pregnancy are the biochemical PAPP-A and β-hCG and the ultrasonographic markers, DV, and the absence or presence of the fetal NB.

TABLE I
EUPLOID AND ANEUPLOID POPULATIONS OF THE THREE DATASETS USED

Dataset	Euploid	T21	T18	T13	Triploidy	Turner
A	50 517	408	39	14	10	13
B	29 790	124	42	10	14	19
C	42 055	152	60	22	14	26

B. Artificial Neural Networks

The major advantage of the use of ANNs compared to other statistical approaches for classification tasks is their ability to learn by examples. A typical architecture of an ANN has one input layer, one or more hidden layers and one output layer. Each layer has a number of nodes that are connected to each other via a weight that represents the synapse in the biological neural networks. The first layer consists equal nodes to the number of input markers. The number of nodes in the hidden layers is a parameter and it has to be optimized manually, according to the problem under study. The last layer (output layer) contains one or more nodes, depending on the number of classes. In tasks with two classes, it is commonly used one node. The value of the output layer is finally passed through a step function where a cutoff point binarizes the output into 0 or 1.

In the training phase of an ANN, all the examples are presented to the nodes of the input layer one by one. The values of the input pattern are first multiplied with the respective weight. Then, all the products between the input values and the weights of every node are summed and passed to the nodes of the first hidden layer, through an activation function. The information from the hidden layer to the next hidden layers and the output layer is passed in a similar way as from the input layer to the hidden layer. One epoch is considered when all the examples are seen by the network and processed through the hidden layer(s) and the output layer. After every epoch, the weights of the ANN are updated based on a cost function that is calculated as an error function between the known target and the output value of the ANN. The size of the weight change that is made in every epoch is controlled by a parameter called "learning rate". One common error function that is used in feed-forward backpropagation networks is the mean squared error (MSE). The training of an ANN converges when the MSE is below a certain threshold, or when the specified epochs are reached. In our experiments, we used feed-forward backpropagation networks with one hidden layer and one node in the output layer. The weights are initialized randomly and the learning rate was set at 0.3. We used 500 epochs for training the ANNs.

C. Cross Validation

The three datasets were split in two sets each of 70% and 30%. The first set is used for training the ANNs. The second set

TABLE II
CROSS VALIDATION

	Training		Validation		
Dataset	Euploid	T21	Euploid	T21	OCA
A	33 619	279	16 898	129	76
B	20 782	87	9 008	37	85
C	31 225	100	10 830	52	122

Training and validation sets for the three datasets used.

TABLE III
SHORT AND LONG GROUPS OF INPUT MARKERS THAT ARE USED AS INPUTS TO THE NEURAL NETWORK. THE ABBREVIATIONS MA, CRL, NT, DV, TF, AND NB STAND FOR MATERNAL AGE, CROWN RUMP LENGTH, NUCHAL TRANSLUCENCY, DUCTUS VENOSUS, TRICUSPID FLOW, AND NASAL BONE, RESPECTIVELY. THE WORD YES INDICATES THAT THE SPECIFIC MAKER IS USED IN THE RESPECTIVE GROUP. SIMILARLY, THE WORD NO INDICATES THAT A MARKER IS NOT USED.

	Long				Short			
Markers	9	8a	7	8b	6	5a	5b	4
MA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CRL	Yes	Yes	No	No	Yes	Yes	No	No
PAPP-A	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
β-hCG	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prev. T21	Yes	No	No	Yes	Yes	No	No	Yes
NT	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV	Yes	Yes	Yes	Yes	No	No	No	No
TF	Yes	Yes	Yes	Yes	No	No	No	No
NB	Yes	Yes	Yes	Yes	No	No	No	No

is kept away from the training procedure as the validation set. The number of cases in each training and validation sets for the three datasets used are shown in Table II.

D. Marker Selection

After the consultation of the doctors that are involved in this research, we have made experiments with the aim of minimizing the required number of prenatal examinations that a pregnant woman is requested to perform. Two groups of markers namely "short" and "long" were examined as a potential two-stage screening for aneuploidy. The "short" group is consisted of the maternal age, the biochemical β-hCG and PAPP-A, and the fetal NT. In the "long" group, we included three additional markers, the DV, the TF, and the absence or presence of the NB that are extracted from an additional special ultrasound examination.

From the two groups of markers, we created eight combinations that are used as input to the ANN. In Table III, we present the combinations used in both "short" and "long" groups. In the second row, we show the ID of every network that corresponds to the figures in Section III.

E. Data Normalization

The conversion of the biochemical markers PAPP-A and β-hCG into their MoMs, and the NT into delta NT has been proposed by Kagan *et al.* [4] as a step in their methodology for the patient-specific risk for T21. Kagan *et al.* [4] show that the values of the biochemical markers and the NT are correlated with the gestational age at the time of the examination.

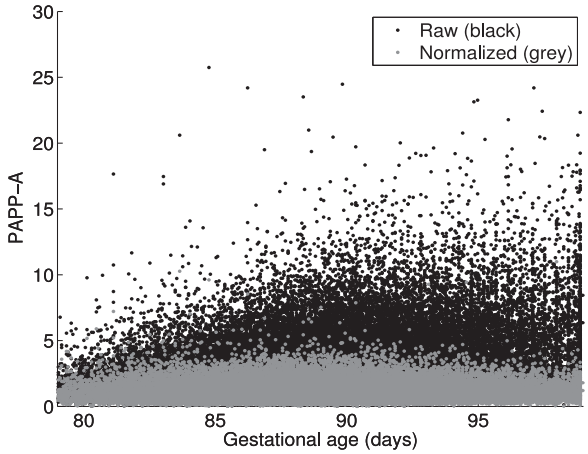

Fig. 2. Raw and normalized (MoM) values of the PAPP-A of all the cases that consist the *Dataset A*. The values are sorted in ascending order based on the gestational age. It is shown that the raw values of the PAPP-A increase with gestational age, while the normalized values are distributed around the average value of 2.5.

In Fig. 2, we superimpose the raw (black dots) and the MoM (gray dots) values for the biochemical marker PAPP-A of all the cases of *Dataset A*. We note that the data were first sorted based on the gestational age. In *X*-axis, we plot the gestational age in days to better visualize the effect that the raw values of PAPP-A are correlated with gestational age. It is shown in Fig. 2 that the values after normalization are distributed in a lower range and the correlation to the gestational age is lost.

F. Data Reduction

The datasets used in this study are highly imbalanced: The euploid cases occupy more than 99% of the total population. In machine learning, the use of imbalanced populations for training may cause several technical problems. For instance, Bayesian classification uses an *a-priori* risk that is based on the population of each class. ANNs adjust their weights according to the MSE of each epoch, as explained in Section II-B. Since the MSE is global for both minority and majority classes, the false predictions of the minority class are not influencing significantly the total MSE. In Fig. 3, we present the MSE of two neural networks trained with 1) imbalanced (solid line) and 2) balanced (dashed line) sets for 500 epochs. It is shown that the MSE of the balanced training set is significantly lower.

Several approaches for creating balanced training sets are proposed in the literature. One way is to choose representative instances from the majority class to reduce the population for training. Another way is to artificially create data to increase the population of the minority class. Creating artificial cases for medical data is a difficult task due to the unpredictable correlation of the markers. For instance, the CRL and the NT are correlated with the gestational age at the time of the examination. This relationship is not fully understood and we are uncertain

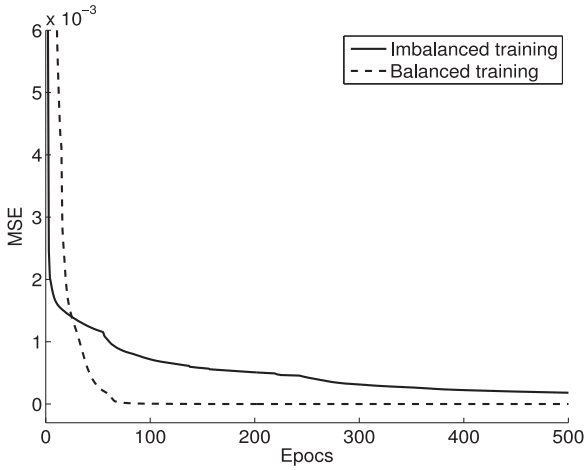

Fig. 3. MSE of the neural network that was trained for 500 epochs. Dashed and solid lines show the MSE of balanced and imbalanced training sets.

TABLE IV
CLUSTER MAP FROM THE *k*-MEANS ALGORITHM

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Input	5 084	15 479	1 619	227	11 210
Reduced	76	231	25	4	167

if we can model artificial data that will follow these specific pattern relations.

There is no "best method" to apply in general because the method is problem dependent and highly relevant to the nature and the complexity of the data under study. This finding is supported also by the fact that ANNs learn by training and analysis does not assume normal distribution.

In this study, we choose the first approach to reduce the population of the majority class. First, we apply unsupervised clustering to the euploid cases using the *k*-means algorithm with five prototypes. The number of the prototypes was selected by applying the elbow method [25]. Then, for every cluster we identify the *k*-nearest neighbors of the respective prototype. The number of k is defined automatically and it is proportional to the length of the respective cluster with the length of the total euploid population:

$$k = \text{target\_population}\frac{\text{size}(\text{cluster}_k)}{\text{size}(\text{euploid})}. \quad (1)$$

In the second row of Table IV, we present the distribution of the *k*-means outputs that were built with the combination of the biomarkers and the ultrasonographic markers. In this example, we choose to reduce the euploid for training from 29 790 to 503 cases. In the last row of Table IV, we present the number of the representative cases that are collected from every cluster, using (1).

In Fig. 4, we present a two-dimensional (2-D) plot of the biochemical markers β-hCG (x-axis) and PAPP-A (y-axis) for the 5 084 cases of the first cluster that are shown with black dots. The prototype of the cluster is shown with a white star and the 76-nearest to the prototype are shown with gray dots. It is emphasized that in this representation, the cases that are chosen as representatives are not necessarily the ones that are closer to the prototype. This is due to the contribution of the other markers used in the training of the *k*-means. That would be the case if *k*-means were trained solely with the two biomarkers β-hCG and PAPP-A.

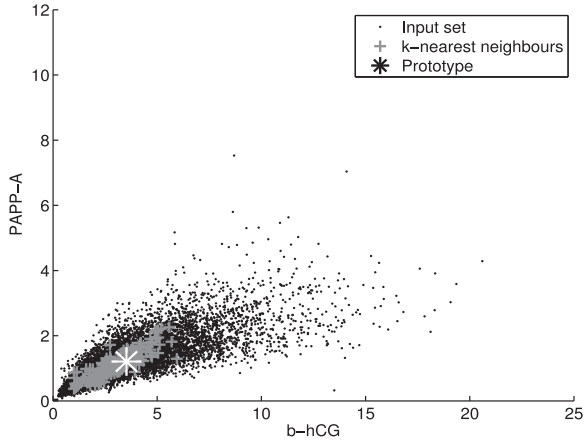

Fig. 4. 2D plot of the biomarkers β-hCG and PAPP-A. The euploid cases of the first k-means cluster are presented with black dots. The prototype is shown with a white star and the 76-nearest to the prototype are shown with gray dots.

G. Evaluation Protocol

Choosing the best ANN architecture for the problem under study is not an easy task and it is usually done empirically by the system designer [17]. In an attempt to choose the optimal neural network architecture and parameters, we followed a grid search approach by using the training set of *Dataset A* to construct 24 neural networks by changing the number of neurons in the hidden layer from 5 to 60, with step of 5 neurons, the logistic, and the hyperbolic tangent transfer functions. The network that returned the best results on the validation set of *Dataset A* was used to construct the networks for the *Datasets B* and *C* that were used in all of our experiments. We use 50 nodes in the hidden layer with logistic activation function.

III. RESULTS

The findings of our experiments are summarized in Figs. 5–7 and Tables V–VII. In Fig. 5, we show the results of the *Dataset B* and we present the FPRs at 100% DR for T21 of the eight ANN models that were built by using different combinations of the input markers. Additionally, we visualize the difference of the results between the networks trained with normalized and raw data with solid and dashed lines, respectively. The difference between the performance of the raw and normalized ANNs is statistically significant for *Dataset B* but not for *Datasets A* and *C*. In *Dataset B*, as shown in Fig. 5, the "short" marker group (ID: 4, 5a, 5b, and 6) returned relatively high FPR ($> 20\%$), whereas the long group (ID: 7, 8a, 8b, and 9) yielded significantly lower FPR.

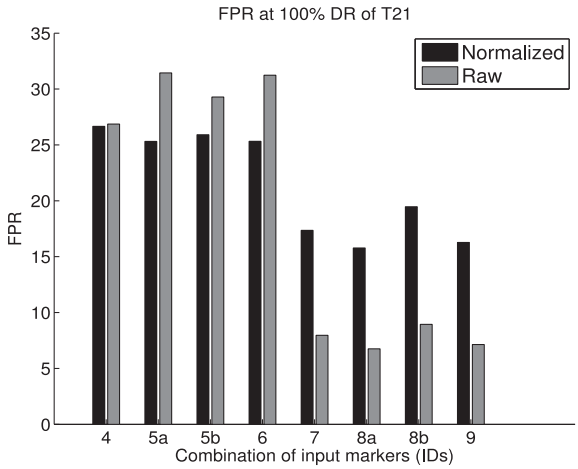

Fig. 5. FPRs of the ANNs built with normalized (black bars) and raw (gray bars) values of *Dataset B* at a T21 DR of 100%. The results of the different input markers combinations are shown in the *x*-axis.

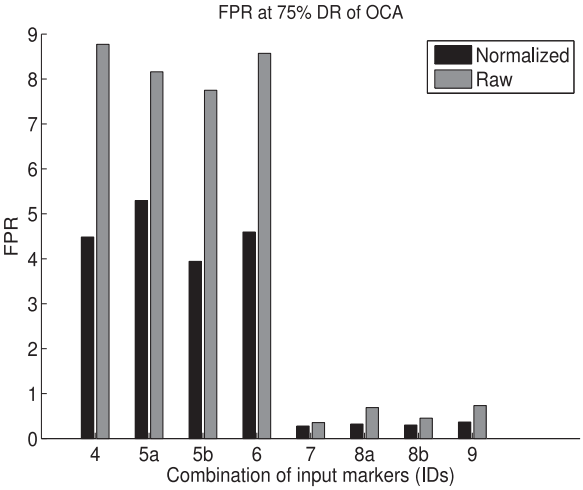

Fig. 6. FPRs of the ANNs built with normalized (black bars) and raw (gray bars) values of *Dataset B* at an OCA DR of 75%. The results of the different input markers combinations are shown in the *x*-axis.

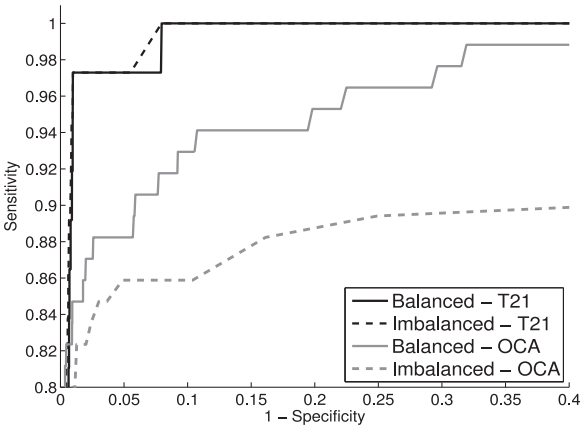

Fig. 7. ROC curves of the models build with balanced (solid lines) and imbalanced (dashed lines) datasets for the euploid-T21 (black lines) and euploid-OCA (gray lines). The data are referred to the seven markers group and raw data.

TABLE V
FPR AND DR FOR T21 AND OCA FOR THE NETWORKS BUILT WITH *Dataset A*

Dataset A	FPR	T21	OCA
Imbalanced—Normalized	2.7%	100.0%	7.9%
Imbalanced—Raw	7.9%	100.0%	32.9%
Balanced—Normalized	10.0%	100.0%	73.7%
Balanced—Raw	8.3%	100.0%	69.7%

The data are referred to the seven markers group and raw data.

TABLE VI
FPR AND DR FOR T21 AND OCA FOR THE NETWORKS BUILT WITH *Dataset B*

Dataset B	FPR	T21	OCA
Imbalanced—Normalized	34.2%	100.0%	90.6%
Imbalanced—Raw	7.9%	100.0%	83.5%
Balanced—Normalized	17.4%	100.0%	94.1%
Balanced—Raw	8.0%	100.0%	89.4%

The data are referred to the seven markers group and raw data.

TABLE VII
FPR AND DR FOR T21 AND OCA FOR THE NETWORKS BUILT WITH *Dataset C*

Dataset C	FPR	T21	OCA
Imbalanced—Normalized	4.1%	100.0%	85.2%
Imbalanced—Raw	4.3%	98.1%	84.4%
Balanced—Normalized	5.8%	100.0%	85.2%
Balanced—Raw	4.8%	100.0%	85.2%

The data are referred to the seven markers group and raw data.

In Fig. 6, we present the FPRs of the same models as in Fig. 5 at the 75% DR of the OCA. The "short" group returned a FPR of 4.5% (in average of four ANNs) and of 8.5% for the models built with the normalized and raw values, respectively. There is no significant difference between the FPRs (average of 0.2%) of the "long" normalized/raw marker groups.

In an attempt to visualize the difference of the performance between the balanced and imbalanced training sets, we calculate the sensitivity and the 1-Specificity for different cutoff values as explained in Section II-B and we present the results of the network built with raw values and the marker set with seven inputs in a receiver operating characteristic (ROC) curve (see Fig. 7). The sensitivity and the specificity are defined as shown in (2) and (3), respectively

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

where TP, FN, TN, and FP are abbreviations of the true positive, false negative, true negative, and false positive, respectively.

In Fig. 7, we superimpose the ROC curves for the networks built with raw marker values for imbalanced and balanced training sets by using the "long" feature set with seven inputs. The results of the balanced and imbalanced training sets are distinguished with solid (black for T21) and dashed lines, respectively (gray for OCA). It is shown that there is a significant difference between the DRs of the OCA built with balanced and imbalanced datasets. The DR of the T21 has no statistical difference.

In Tables V–VII, we show the results of the networks built with the balanced and imbalanced training sets and the raw and the normalized seven input markers for the three datasets. In the first column of every table, we present the type of the network (balanced or imbalanced training sets, and raw or normalized markers). In the second column, we show the FPR and in the last two the DRs for the T21 and the OCA, respectively.

The best results of the experiments done with *Datasets B* and *C* were achieved with the ANNs built with the seven markers group and raw balanced data (8% and 4.8% FPR). The performance of the ANN built with the normalized *Dataset B* is significantly weaker with a difference more than 10% of FPR, compared to the ANNs built with the raw data. The same experiment was done with *Dataset C* with no difference on the results between raw and normalized data. With the results of *Dataset B*, we suggest that the normalization of the data could be avoided.

IV. Discussion

In this paper, we demonstrate the effectiveness of the ANNs schemes as a potential classifier for the diagnosis of the T21 and OCA, in the early stage of a pregnancy. Our principle aim is to build models that ensure false negative classifications of T21 at the lowest possible rate. The best results were achieved with the seven markers group (*Dataset A*: 100% DR of T21 and FPR of 2.7%) with normalized imbalanced data. However, the *Dataset A* was used as a training set to perform a grid search over several parameters and architectures of the ANNs, as explained in Section II-G. The same structure of ANN was used in *Datasets B* and *C* and we found that the results are not consistent with *Dataset A*. Additionally, from Table V, we observe that the abovementioned ANN built with *Dataset A* returns the lowest DR of the OCA.

Another objective of this study was to optimize the FPR with respect to the cost of every examination that is necessary to be done for the estimation of the risk for aneuploidy. In principle, the task is to determine the optimum number of markers required as input to the classifier. We have examined the robustness of the ANNs that are built with different combinations of input markers and we suggest two groups namely the "short" and the "long." The "short" group consisted of markers that can be extracted in one visit to the doctor. We achieve 100% DR of T21 at a relatively high FPR of 25%. The "long" markers group consists of three additional markers that can be extracted in another examination that measures the flow of the DV and the TF by using a Doppler technique. The third marker is the NB that can be visualized during the ultrascan. The "long" markers group achieves a lower FPRs of 5%, at the same DR of 100% for T21 and $>$80% DR of OCA.

The population under study is not normally distributed by nature since pregnancy takes place in a certain age range, which is by nature skew to the right. In addition, the large database used in this study is highly imbalanced due to the very low prevalence of aneuploidy cases in the general population, even though it contains much more Trisomy cases than reported statistically. For example, T21 occurs 1 in 800 pregnancies, and thus in our database, we should have had less T21 cases than we actually have. Similarly, T18 occurs 1 in 5 000 pregnancies and, thus, we should again have even less cases.

One other possible reason of the nonnormality distributions of the euploid may be due to the increased significance of those values that depart from normality. This is usually a case when there is a large population such as the population of the euploid cases in our database. Any method that measures the normality of a distribution of large populations has high probability of rejecting the null hypothesis that the sample comes from a normal distribution.

From our experiments, see Fig. 5 in Section III, we conclude that marker CRL does not contribute significantly to the diagnostics of the system and hence it can be ignored. The CRL is a physical distance between the crown and the rump of the fetus and the obstetrician measures it during the ultrascan. An accurate measurement of the CRL requires a specific position of the fetus and other factors that are unpredictable. Due to this fact, the distribution of the CRL values has a high standard deviation.

The balanced ANNs shown in Fig. 7 yield lower FPRs compared to imbalanced ANNs for the detection of the OCA, while no significant difference was found for the detection of T21. Nevertheless, we suggest the use of a balanced training set for the ANNs as the MSE reduces dramatically compared to the score of the imbalanced training set. From our results, we demonstrate that the networks trained with balanced populations of euploid and aneuploid yield lower FPRs compared to the imbalanced training sets and, therefore, we suggest this method as a preprocessing step.

The estimation of the risk for T21 is currently done by using an SMM from the combination of the markers that are explained in this paper and there is a DR of 95% of T21 at 5% FPR. Our method outperforms the state-of-the-art method with 100% DR of the T21 at the same FPR. Additionally, our method detects $>$80% of the OCA. From the two other noninvasive methods (SSM and cell-free fetal DNA) for the diagnosis of aneuploidy that are found in the literature, the cell-free fetal DNA test is the most promising. Several published studies report perfect separation between the euploid and aneuploid. However, there is a practical problem of this method being the cost, which is forbidding for general use. Moreover, the results are returned after some considerable time. The proposed methodology demonstrated in this paper can be used as a first screening on the data for selecting the positively ranked cases (100% T21, 85% OCA at 5% FPR), which will be the only ones suggested for a cell-free fetal DNA test. This will limit the overall cost for prenatal test and at the same time guarantee zero undiagnosed T21 births.

V. Future Work

The combination of the "short" and "long" markers groups could be a two-stage procedure for a first and second screening for aneuploidy. The results of the "short" marker group are returned immediately in the personal computer of the doctor in one visit. These results assure that the negative prediction does not contain any T21. All the positive cases will be reevaluated with the "long" group to estimate the final risk for T21 and OCA. This paper will be validated and reported in the future work.

VI. Conclusion

Diagnosis of the T21 and OCA can be effectively achieved with ANNs and a combination of biomarkers and ultrasonographic markers in the early stage of pregnancy. We have used three datasets to answer our research questions that include the detection of all the T21 cases at the lowest FPR possible, the identification of an optimal combination of input markers, the contribution of the normalized values of the data over the raw and the possible use of training sets that are consisted with balanced populations among euploid and aneuploid.

We have shown that the optimal combination of markers belongs to the "long group" that requires ultrasonographic and maternal blood examinations. Furthermore, the use of the raw data appear to be significantly more effective for the networks build with the "long group" but less effective for the "short group." Another contribution of this paper is the proposed method for the data reduction of the euploid by using the *k*-means algorithm in an attempt to create populations balanced in numbers among euploid and aneuploid. We have shown that the balanced sets appear to be more effective for training the ANNs.

In this paper, we present a system that is able to identify the entire population of T21 and the majority of the OCA, such as T18, T13, Turner syndrome, and Triploidy. We have used datasets with populations that ensure statistical confidence of our results, compared to databases used in similar work of other groups. We achieved with ANNs a 100.0% DR of T21 and 85.2% of the OCA with FPR of 4.8%.

Acknowledgment

The authors would like to thank Dr. C. Neocleous and Dr. K. Neokleous from the CI Lab of the University of Cyprus for their valuable assistance during the development of the ANN models, Dr. A. Syngelaki from FMF for preparing and managing the database used in this study and Prof. N. Petkov for reviewing the paper and his valuable guidance.

References

[1] C. Enzensberger *et al.*, "Fetal loss rate and associated risk factors after amniocentesis, chorionic villus sampling and fetal blood sampling," *Ultraschall Medizin*, vol. 33, no. 7, Dec. 2012, Art. no. E75-9.

[2] R. J. M. Snijders, K. Sundberg, W. Holzgreve, G. Henry, and K. H. Nicolaides, "Maternal age- and gestation-specific risk for trisomy 21," *Ultrasound Obstetrics Gynecology*, vol. 13, pp. 167–170, Dec. 1999.

[3] R. J. M. Snijders *et al.*,, "UK multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal-translucency thickness at 10–14 weeks of gestation," *Lancet*, vol. 352, no. 9125, pp. 343–346, 1998.

[4] K. Kagan, D. Wright, A. Baker, D. Sahota, and K. Nicolaides, "Screening for trisomy 21 by maternal age, fetal nuchal translucency thickness, free beta-human chorionic gonadotropin and pregnancy-associated plasma protein-A," *Ultrasound Obstetrics Gynecology*, vol. 31, no. 6, pp. 618–624, 2008.

[5] K. Kagan, D. Wright, K. Spencer, F. Molina, and K. Nicolaides, "First-trimester screening for trisomy 21 by free beta-human chorionic gonadotropin and pregnancy-associated plasma protein-A: Impact of maternal and pregnancy characteristics," *Ultrasound Obstetrics Gynecology*, vol. 31, no. 5, pp. 493–502, 2008.

[6] K. Spencer, V. Souter, N. Tul, R. Snijders, and K. Nicolaides, "A screening program for trisomy 21 at 10–14 weeks using fetal nuchal translucency, maternal serum free β-human chorionic gonadotropin and pregnancy-associated plasma protein-A," *Ultrasound Obstetrics Gynecology*, vol. 13, no. 4, pp. 231–237, 1999.

[7] K. Spencer, C. E. Spencer, M. Power, C. Dawson, and K. H. Nicolaides, "Screening for chromosomal abnormalities in the first trimester using ultrasound and maternal serum biochemistry in a one-stop clinic: A review of three years prospective experience," *BJOG: Int. J. Obstetrics Gynaecology*, vol. 110, no. 3, pp. 281–286, 2003.

[8] K. H. Nicolaides, "First-trimester screening for chromosomal abnormalities," *Semin. Perinatology*, vol. 29, no. 4, pp. 190–194, 2005.

[9] N. Maiz, C. Valencia, K. Kagan, D. Wright, and K. Nicolaides, "Ductus venosus Doppler in screening for trisomies 21, 18 and 13 and turner syndrome at 11–13 weeks of gestation," *Ultrasound Obstetrics Gynecology*, vol. 33, no. 5, pp. 512–517, 2009.

[10] I. Huggon, D. DeFigueiredo, and L. Allan, "Tricuspid regurgitation in the diagnosis of chromosomal anomalies in the fetus at 11–14 weeks of gestation," *Heart*, vol. 89, no. 9, pp. 1071–1073, 2003.

[11] G. Ashoor, A. Syngelaki, M. Wagner, C. Birdir, and K. H. Nicolaides, "Chromosome-selective sequencing of maternal plasma cell-free DNA for first-trimester detection of trisomy 21 and trisomy 18," *Amer. J. Obstetrics Gynecology*, vol. 206, no. 4, 2012, Art. no. 322.e1-5.

[12] E. A. Papageorgiou, A. Karagrigoriou, E. Tsaliki, V. Velissariou, N. P. Carter, and P. C. Patsalis, "Fetal-specific DNA methylation ratio permits noninvasive prenatal diagnosis of trisomy 21," *Nature Med.*, vol. 17, no. 4, pp. 510–513, 2011.

[13] H. Zhang *et al.*, "Non-invasive prenatal testing for trisomies 21, 18 and 13: Clinical experience from 146 958 pregnancies," *Ultrasound Obstetrics Gynecology*, vol. 45, no. 5, pp. 530–538, 2015.

[14] T. Sahoo *et al.*, "Prenatal diagnosis of chromosomal abnormalities using array-based comparative genomic hybridization," *Genetics Med.*, vol. 8, no. 11, pp. 719–727, 2006.

[15] K. O. Kagan, D. Wright, and K. Nicolaides, "First-trimester contingent screening for trisomies 21, 18 and 13 by fetal nuchal translucency and ductus venosus flow and maternal blood cell-free DNA testing," *Ultrasound Obstetrics Gynecology*, vol. 45, no. 1, pp. 42–47, 2015.

[16] D. S. Levine, *Introduction to Neural and Cognitive Modeling*. Hove, U.K.: Psychology Press, 2000.

[17] A. Neocleous, K. Nicolaides, and C. Schizas, "First trimester non-invasive prenatal diagnosis: A computational intelligence approach," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 5, pp. 1427–1438, Sep. 2016.

[18] F. Schnorrenberg, C. S. Pattichis, K. C. Kyriacou, and C. N. Schizas, "Computer-aided detection of breast cancer nuclei," *IEEE Trans. Inf. Technol. Biomed.*, vol. 1, no. 2, pp. 128–140, Jun. 1997.

[19] C. N. Schizas and C. S. Pattichis, "Learning systems in biosignal analysis," *BioSystems*, vol. 41, no. 2, pp. 105–125, 1997.

[20] J. Neves *et al.*, "Logic programming and artificial neural networks in breast cancer detection," in *Advances in Computational Intelligence*, New York, NY, USA: Springer-Verlag, 2015, pp. 211–224.

[21] O. Er, O. Cetin, M. S. Bascil, and F. Temurtas, "A comparative study on Parkinson's disease diagnosis using neural networks and artificial immune system," *J. Med. Imag. Health Informat.*, vol. 6, no. 1, pp. 264–268, 2016.

[22] Y. Wang, J. Li, J. Gu, Z. Zhou, and Z. Wang, "Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China)," *Appl. Soft Comput.*, vol. 35, pp. 280–290, 2015.

[23] T. K. Yoo, D. W. Kim, S. B. Choi, E. Oh, and J. S. Park, "Simple scoring system and artificial neural network for knee osteoarthritis risk prediction: A cross-sectional study," *PLoS One*, vol. 11, no. 2, 2016, Art. no. e0148724.

[24] Q. Liu, X. Cui, Y.-C. Chou, M. F. Abbod, J. Lin, and J.-S. Shieh, "Ensemble artificial neural networks applied to predict the key risk factors of hip bone fracture for elders," *Biomed. Signal Process. Control*, vol. 21, pp. 146–156, 2015.

[25] R. L. Thorndike, "Who belongs in the family?" *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1953.

Andreas C. Neocleous was born in Cyprus. He received the degree in audio signal processing and acoustics from the Technical University of Crete, Heraclion, in 2009, and the Graduate in sound and music computing in 2010 from the University of Pompeu Fabra, Barcelona, Spain. He is currently working toward the Ph.D. degree at the University of Groningen, Groningen, The Netherlands.

Since 2011, he has been collaborating with the University of Cyprus (UCY) as a Research Scientist on research programs funded by the EU, the UCY, and the Cyprus Research Promotion Foundation. He has published articles and presented his work at international conferences. His research interests include digital signal processing, machine learning, and computational intelligence applied in medical diagnostic systems.

Kypros H. Nicolaides was born in Cyprus. He received the degree in medicine from King's College London, London, U.K. He trained in obstetrics and gynecology and specialized in fetal medicine in 1980.

He is currently a Professor of fetal medicine at King's College London. He has carried out extensive research in many aspects of fetal diagnosis and therapy leading to the publication of more than 1200 peer-reviewed papers in scientific journals. He provided training in fetal medicine to more than 500 doctors from 50 countries and has supervised more than 50 doctors to undertake research leading to Ph.Ds and M.Ds. In 1995, he founded the charity Fetal Medicine Foundation, which has donated more than 15 million to promote research and training in Fetal Medicine throughout the World.

Prof. Nicolaides received the following awards: the Ian Donald Gold Award for highest contribution in ultrasound of the International Society Ultrasound in Obstetrics and Gynecology, in 1999; the Eric Saling Award for highest scientific contribution in perinatal medicine of the World Association of Perinatal Medicine, in 2001; Excellence in Letters, Culture and Science, Government of Cyprus, in 2004; the Honorary Fellowship of the American Institute of Ultrasound in Medicine, USA, in 2004; and the Honorary Doctorate in Medicine: Universities of Athens and Ioannina in Greece, University of Warsaw in Poland, University of Bucharest in Romania, University of Jinan in China, University of Amsterdam in the Netherlands, University of Olomouc in the Czechia, European University of Cyprus.

Christos N. Schizas (SM'81) was born in Cyprus. He received the B.Sc. (Eng.) degree in electronics from Queen Mary College, University of London, London, U.K., in 1978, the M.B.A. degree in business administration from the University of Indianapolis, Indianapolis, IN, USA, in 1988, and the Ph.D. degree in intelligent systems from Queen Mary College, University of London, in 1981.

He is currently a Professor of computational intelligence at the University of Cyprus, Nicosia, Cyprus. He was a Postdoctoral Fellow with the University of London, and a Professor of computer information systems with the University of Indianapolis. Since 1991, he has been with the Department of Computer Science, University of Cyprus. He was the Vice Rector of the University during 2002–2006. His research interests include eHealth, computational intelligence, medical informatics, diagnostic and prognostic systems, and system modeling and identification of brain activity.

Dr. Schizas received the William Lincoln Shelley Award from the University of London for excellence in research, and a Fulbright fellowship for collaborative research in the USA. He is currently the Section Editor (*eHealth*) of *Technology and Health Care*, served as an Area Editor of the IEEE TRANSACTIONS OF INFORMATION TECHNOLOGY IN BIOMEDICINE, and a Member of the Editorial Board of the *Journal of Intelligent Systems*. He is the Founder of the Computational Intelligence Lab and the Co-Director of the eHealth lab of the University of Cyprus. He has taken part in European Commission initiatives for promoting the Information Society, especially the Euro-Mediterranean partnership and the eHealth initiatives. He attends regularly as Invited Speaker at the annual EU Ministerial Forum week meeting on *eHealth*. In April 2016 he has been appointed by the President of the Republic of Cyprus as Presidential Counselor for *eHealth* and matters related to the national health system reform.