



Artificial intelligence in prenatal diagnosis: Down syndrome risk assessment with the power of gradient boosting-based machine learning algorithms

Prenatal tanıda yapay zeka: Gradient boosting tabanlı makine öğrenmesi algortimalarının gücü ile Down sendromu risk değerlendirmesi

Emre Yalçın¹, Tarık Kaan Koç², Serpil Aslan², Süleyman Cansun Demir¹, İsmail Cüneyt Evrûke¹, Mete Sucu¹, Mesut Avan¹, Fatma İşlek Uzun¹

¹Çukurova University Faculty of Medicine, Department of Obstetrics and Gynecology, Division of Perinatology, Adana, Türkiye

²Malatya Turgut Özal University Faculty of Engineering and Natural Sciences, Department of Software Engineering, Malatya, Türkiye

Abstract

Objective: One of the most common chromosomal abnormalities seen during pregnancy is Down syndrome (Trisomy 21). To determine the risk of Down syndrome, first-trimester combined screening tests are essential. Using data from the first-trimester screening test, this study compares machine learning and deep learning models to forecast the risk of Down syndrome.

Materials and Methods: Within the scope of the study, biochemical and biophysical data of 959 pregnant women who underwent first-trimester screening tests at Çukurova University Obstetrics and Gynecology Clinic between 2020-2024 were analyzed. After cleaning missing and erroneous data, various preprocessing and normalization techniques were applied to the final dataset consisting of 853 observations. Down syndrome risk prediction was performed using different machine learning models, and model performances were compared based on accuracy rates and other evaluation metrics.

Results: Experimental results show that the CatBoost model provides the highest success rate, with an accuracy rate of 95.31%. In addition, the XGBoost and LightGBM models exhibited high performance, with accuracy rates of 95.19% and 94.84%, respectively. The study also examines the effects of the class imbalance problem on model performance in detail and evaluates various strategies to reduce this imbalance.

Conclusion: The findings show that gradient boosting-based machine learning models have significant potential in Down syndrome risk prediction. This approach is expected to contribute to the reduction of unnecessary invasive tests and improve clinical decision-making processes by increasing the accuracy rate in prenatal screening processes. Future studies should aim to increase the generalization capacity of the model on larger data sets and to provide integration with different machine learning algorithms.

Keywords: Down syndrome, first-trimester screening test, gradient boosting, machine learning, artificial intelligence, classification algorithms

PRECIS: The aim of the study is to increase the accuracy rate in prenatal screening processes, and it is expected to contribute to reducing unnecessary invasive tests and improving clinical decision-making processes.

Corresponding Author/Sorumlu Yazar: Emre Yalçın PhD,

Çukurova University Faculty of Medicine, Department of Obstetrics and Gynecology, Division of Perinatology, Adana, Türkiye

E-mail: dremreyalcin23@gmail.com ORCID ID: orcid.org/0000-0001-7088-6749

Received/Geliş Tarihi: 15.02.2025 Accepted/Kabul Tarihi: 26.05.2025 Epub: 27.05.2025

Cite this article as: Yalçın E, Koç TK, Aslan S, Demir SC, Evrûke İC, Sucu M, et al. Artificial intelligence in prenatal diagnosis: Down syndrome risk assessment with the power of gradient boosting-based machine learning algorithms. Turk J Obstet Gynecol. [Epub Ahead of Print]



Öz

Amaç: Down sendromu (Trizomi 21), prenatal dönemde en sık rastlanan kromozomal anomalilerden biridir. Gebeliğin birinci trimesterinde uygulanan kombine tarama testleri, Down sendromu riskinin belirlenmesi için önemli bir araç olarak kullanılmaktadır. Bu çalışma, birinci trimester tarama testi verileri kullanılarak Down sendromu riskini tahmin etmek amacıyla farklı makine öğrenmesi ve derin öğrenme modellerini karşılaştırmalı olarak değerlendirmeyi amaçlamaktadır.

Gereç ve Yöntemler: Çalışma kapsamında, 2020-2024 yılları arasında Çukurova Üniversitesi Kadın Doğum Kliniği'nde birinci trimester tarama testine tabi tutulan 959 gebeye ait biyokimyasal ve biyofiziksel verileri analiz edilmiştir. Eksik ve hatalı veriler temizlendikten sonra, 853 gözlemden oluşan nihai veri seti üzerinde çeşitli ön işleme ve normalizasyon teknikleri uygulanmıştır. Farklı makine öğrenmesi modelleri kullanılarak Down sendromu risk tahmini gerçekleştirilmiş, model performansları doğruluk oranları ve diğer değerlendirme metrikleri üzerinden karşılaştırılmıştır.

Bulgular: Deneyisel sonuçlar, CatBoost modelinin %95,31 doğruluk oranı ile en yüksek başarıyı sağladığını göstermiştir. Bunun yanı sıra, XGBoost ve LightGBM modelleri sırasıyla %95,19 ve %94,84 doğruluk oranları ile yüksek performans sergilemiştir. Çalışmada ayrıca sınıf dengesizliği probleminin model performansı üzerindeki etkileri detaylı olarak incelenmiş ve bu dengesizliği azaltmaya yönelik çeşitli stratejiler değerlendirilmiştir.

Sonuç: Elde edilen bulgular, gradient boosting tabanlı makine öğrenmesi modellerinin Down sendromu risk tahmininde önemli bir potansiyele sahip olduğunu göstermektedir. Bu yaklaşımın, prenatal tarama süreçlerindeki doğruluk oranını artırarak, gereksiz invaziv testlerin azaltılmasına ve klinik karar alma süreçlerinin iyileştirilmesine katkı sağlaması beklenmektedir. Gelecekteki çalışmalar, daha geniş veri setleri üzerinde modelin genelleştirme kapasitesini artırmayı ve farklı makine öğrenmesi algoritmalarıyla entegrasyon sağlamayı hedeflemelidir.

Anahtar Kelimeler: Down sendromu, ilk trimester tarama testi, gradyan güçlendirme, makine öğrenmesi, yapay zeka, sınıflandırma algoritmaları

Introduction

Down syndrome (DS) is one of the most common chromosomal abnormalities in humans. It affects individuals regardless of race, age, or socioeconomic status. The condition occurs due to a genetic anomaly in which an extra chromosome is present in the 21st pair, resulting in a total of 47 chromosomes. The incidence rate is estimated to be approximately 1 in every 600 to 800 live births^(1,2). DS is associated with various physical, cognitive, and developmental challenges, along with a range of health complications. Studies have shown that increasing maternal age significantly elevates the risk of DS. Nevertheless, early diagnosis and appropriate management are achievable through prenatal screening tests and genetic counseling.

First-trimester screening (FTS) is a fundamental method for the early detection of DS. This approach integrates maternal serum biomarkers, including free beta-human chorionic gonadotropin (β -hCG) and pregnancy-associated plasma protein A (PAPP-A), with ultrasound-based parameters such as nuchal translucency (NT), crown-rump length (CRL), and the absence or presence of the nasal bone. NT measurements are typically performed between the 11th and 14th weeks of gestation. In cases of DS, β -hCG levels are often elevated, whereas PAPP-A levels tend to be reduced. In contrast, Trisomy 18 and Trisomy 13 are generally associated with lower levels of both markers^(3,4).

Accurate interpretation of screening results requires a clear understanding of the multiple of the median (MoM) method. This approach standardizes test values by dividing each measurement by the median value corresponding to the specific gestational week⁽⁵⁾. Indicators of high risk for DS include a NT measurement greater than 2.5 millimeters, absence of the nasal bone, a PAPP-A level below 0.4 MoM, and a β -hCG level above 2.5 MoM. When NT is 3 millimeters or greater or exceeds the 99th percentile, further fetal evaluation and genetic counseling are strongly recommended. In such cases, additional risk

assessment using cell-free fetal DNA (cfDNA) analysis and confirmatory diagnostic testing should also be considered⁽⁶⁾.

Combined screening tests performed between the 11th and 14th weeks of pregnancy typically yield a false-positive rate of approximately 5 percent and an overall accuracy rate approaching 90 percent. Based on these results, risk levels are categorized as high (equal to or greater than 1 in 250), moderate (between 1 in 250 and 1 in 1000), or low (equal to or less than 1 in 1000)^(7,8). If the screening results are abnormal, amniocentesis is usually recommended. This invasive procedure, typically conducted between the 15th and 20th weeks of gestation, involves extracting fetal cells from the amniotic fluid for genetic analysis⁽⁹⁾. Although it is considered a reliable diagnostic method, amniocentesis carries a small risk of fetal loss. These risks, while infrequent, emphasize the importance of developing more accurate and non-invasive alternatives⁽¹⁰⁾.

In predictive classification, the combination of NT measurements with serum biomarkers enhances the accuracy of DS risk assessment. Artificial intelligence (AI) models are capable of identifying complex patterns within such data, allowing for more precise classification of risk levels⁽¹¹⁾. AI, particularly through machine learning (ML) and deep learning (DL) approaches, facilitates the analysis of large and complex datasets across various disciplines⁽¹²⁾. In the field of healthcare, these technologies have led to faster diagnoses and more efficient treatment planning⁽¹³⁾.

Conventional DS screening methods may be subject to errors due to limitations in clinical expertise or access to advanced technology. In some cases, families may also decline NT measurement because of cultural or personal beliefs. This study integrates both biophysical markers (NT) and biochemical indicators (hCG and PAPP-A) to assess DS risk. The primary objective is to develop a model that reduces the impact of geographic variability and increases the robustness of predictions despite potential test inaccuracies. Data were collected from

959 singleton pregnancies at Çukurova University between 2020 and 2024. After preprocessing, the dataset was used to train AI-based classification models. The outcomes aim to enhance diagnostic accuracy and assist clinicians in prenatal risk evaluation and decision-making.

Materials and Methods

This study effectively estimates the risk of DS, contributing to the health and general well-being of both the mother and the unborn child. The following sections comprehensively explain the applied methodological approaches and present the findings, demonstrating the accuracy and clinical significance of the results.

Dataset and Preprocessing

This study analyzed data obtained from the combined double screening tests of 959 women with singleton pregnancies in the first trimester at the Obstetrics and Gynecology Unit of Çukurova University between 2020 and 2024. The study protocol and data collection process were reviewed and approved by the Çukurova University Faculty of Medicine Research Ethics Committee in accordance with ethical standards (approval number: 144, date: 10.05.2024). Patient records were retrieved from the hospital's gynecology and obstetrics clinic as well as the biochemistry laboratories. To ensure confidentiality and compliance with ethical regulations, all patient data were anonymized, and no personally identifiable information was used at any stage. Details of the dataset and the variables included in the analysis are presented in Table 1.

Prior to the development of the AI model for risk estimation, several preprocessing steps were applied to the dataset to address missing data and improve data quality. Erroneous entries were corrected, and records with duplicate or missing values were excluded. As a result, the initial dataset of 959 records was reduced to 853 valid entries. To improve model accuracy, the distribution of the target variable (DS risk class) was examined. The final dataset included 195 samples (22.9%) classified as medium risk, 474 samples (55.6%) as low risk, and 184 samples (21.6%) as high risk. The categorical distribution of the target variable is illustrated in Figure 1.

Normalization⁽¹⁴⁾

The normalization process was used to improve the model's accuracy and stability because the data set's of the independent variables in the study's data set varying value ranges could cause a scale difference between the variables. Normalization contributes to the stable and efficient operation of ML algorithms by ensuring that the variables are represented on the same scale. Different transformation techniques were evaluated for the target and independent variables in the data preprocessing stage, and appropriate methods were determined.

In particular, since the target variable is categorical, the label encoding method was preferred for the appropriate transformation of the classes. This method allows ML algorithms to process categorical variables more effectively by converting them to numerical values. In the scaling process of the independent variables, three different normalization techniques were tested:

Minimum-Maximum Scaler: This method scales variables to a specific range (usually between 0 and 1) to ensure all features remain within the same limits. Min-max scaling is especially effective when the data is concentrated in a particular range and is preferred when distribution distortion needs to be prevented.

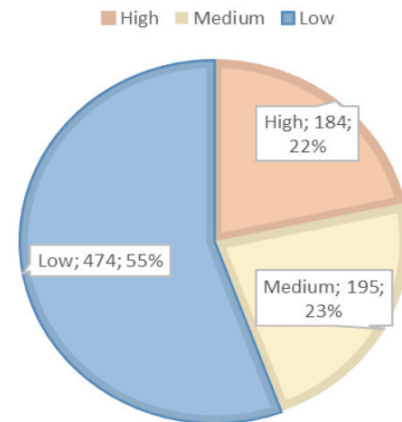


Figure 1. Distribution of the Down syndrome attribute in basic dataset

Table 1. The attributes explanation of the collected dataset

Attribute	Description	Data type
hCG-MoM	Human chorionic gonadotropin (hCG), method of medians (MoM). hCG is a hormone secreted by the placenta during pregnancy and is found at high levels in the early stages of pregnancy. The MoM value is obtained by dividing the measured value by the average value for the corresponding gestational week.	Numeric
PAPP-A MoM	Pregnancy-associated plasma protein A (PAPP-A) is a protein produced by the placenta during pregnancy, known as PAPP-A.	Numeric
NT	Nuchal translucency (NT) is the measurement of the thickness of the fetal neck obtained via ultrasound.	Numeric
NT MoM	It is calculated by dividing the NT value by the average value for the corresponding gestational week.	Numeric
DS	Down syndrome risk classification is as follows: a ratio below 1:250 is considered high risk (high), between 1:250 and 1:1000 is classified as medium risk (medium), and above 1:1000 is categorized as low risk (low).	Categorical

Standard Scaler: This method standardizes the mean of the variables to 0, and the standard deviation to 1, resulting in a standard normal distribution. It is an effective technique, especially for variables with a normal distribution, and is widely used for many ML algorithms.

Robust Scaler: This approach, which applies data scaling based on the median and interquartile range, was designed to minimize sensitivity to outliers. Transforming the data using central tendency measures reduces the negative impact of extreme values on the model. Comparisons conducted within the scope of the study revealed that the Robust Scaler method yielded the most successful results, particularly in cases where outliers were present in the dataset.

As a result of this process, the scaling of the data set was completed, and the model's performance was intended to improve. The normalization process allows the model to learn faster and more stably while improving its prediction performance.

AI Classification Models

In this study, a comprehensive selection of AI-based classification models was utilized to evaluate their predictive performance in assessing DS risk. The models were chosen based on their demonstrated effectiveness in addressing class imbalance, capturing complex non-linear relationships among variables, and performing well in clinical risk classification contexts.

The classifiers were selected with careful consideration of the dataset's characteristics, including its numerical structure, class imbalance, and risk-based output labels. Tree-based ensemble methods such as CatBoost, XGBoost, and LightGBM were included due to their ability to handle structured clinical data effectively. These models are known for their robustness against outliers, high predictive accuracy, and efficient processing in

large datasets. Their successful application in previous prenatal and healthcare-related classification tasks further supports their appropriateness for this study.

An overview of the models and their technical characteristics is provided in Table 2.

Proposed Approach

This study proposes a novel diagnostic approach for early DS risk assessment by integrating biochemical markers (hCG and PAPP-A) and a biophysical parameter (NT) obtained from the combined FTS. These markers were selected based on their well-established roles in prenatal screening. While hCG and PAPP-A provide insights into biochemical deviations associated with chromosomal abnormalities, NT offers a structural sonographic dimension. Combining these complementary features enhances the reliability of early risk estimation.

The novelty of the proposed method lies in the application of advanced AI-based classification models, which go beyond the static threshold-based decisions of traditional screening tools. Unlike conventional methods, the AI-supported approach can capture complex, non-linear interactions between features, enabling more precise and individualized risk stratification. This is particularly important in cases where conventional cut-off values may misclassify borderline or atypical presentations. In addition, the proposed model addresses specific limitations such as operator dependency in NT measurements and potential false reassurance in low-risk cases. By leveraging the learning capabilities of ML algorithms, the model contributes to improving diagnostic robustness and reducing unnecessary invasive procedures.

The general architecture of the proposed methodology is presented in Figure 2.

Table 2. Classification models used in the study and their explanations

Model	Algorithm category	Framework/library	Purpose
Extreme gradient boosting ⁽¹⁵⁾	Gradient boosting	XGBoost	Classification tasks with high performance.
Light gradient boosting machine ⁽¹⁶⁾	Gradient boosting	LightGBM	Efficient gradient boosting for large datasets.
Categorical boosting (CatBoost) ⁽¹⁷⁾	Gradient boosting	CatBoost	Handles categorical features automatically.
K-nearest neighbors (KNN) ⁽¹⁸⁾	Instance-based	scikit-learn	Simple classifier based on distance metrics.
Naive bayes (NB) ⁽¹⁹⁾	Probabilistic	scikit-learn	Based on Bayes' theorem, handles continuous data.
Logistic regression (LR) ⁽²⁰⁾	Linear model	scikit-learn	Binary and multiclass classification.
Support vector machine (SVM) ⁽²¹⁾	Support vector machine	scikit-learn	Classification with margin maximization.
Decision tree (DT) ⁽²²⁾	Tree-based	scikit-learn	Simple, interpretable decision rules.
Random forest (RF) ⁽²³⁾	Ensemble (bagging)	scikit-learn	Combines multiple decision trees for accuracy.
Adaptive boosting (AdaBoost) ⁽²⁴⁾	Ensemble (boosting)	scikit-learn	Boosting technique to improve weak classifiers.
CNN ⁽¹²⁾	Deep learning	TensorFlow/Keras	Pattern recognition in high-dimensional data.
LSTM ⁽¹²⁾	Deep learning (recurrent neural networks)	TensorFlow/Keras	Sequence modeling for time series or text.

Results

The study used ten distinct ML and two DL classifiers to evaluate the suggested strategy's effectiveness. The k-fold cross-validation method ($k=5$) was selected to assess the ML classifiers. Accuracy, Precision, Recall, and F1-score, among the widely used evaluation criteria, were used as the performance scale of the ML classifiers^(12,25).

As can be seen in Figure 3, the highest accuracy performance was obtained with Boosting-based methods. After the comparison, CatBoost obtained the best performance with an accuracy of 95.31%. XGBoost came in second with 95.19%, and LightGBM came in third with 94.84%. In this study, which

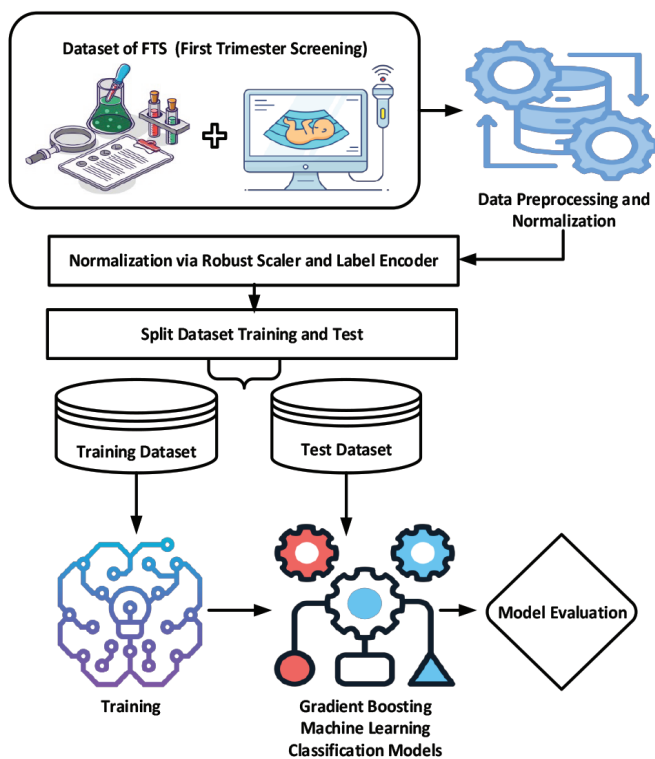


Figure 2. Flowchart of the proposed approach

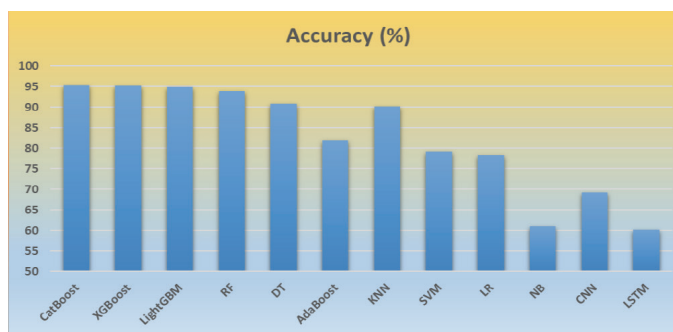


Figure 3. Accuracy performance comparison of classification models

primarily worked with numerical data, tree-based ML models such as CatBoost and XGBoost were preferred due to their robustness against outliers and strong generalization capacity. These models exhibit high performance in learning complex relationships, different features in the dataset, and can capture interactions between variables. In addition, imbalances in the class distribution in the dataset can be managed more effectively thanks to the flexible structure of tree-based models. Such models can make more balanced and adaptable predictions for each class through the use of decision trees to determine patterns in the dataset. In addition, the LightGBM model works with high speed, and low memory usage on large data sets, making it practical to prefer this model.

Because of the unequal distribution of classes, it is important to look at the metrics for each class separately. It is essential to analyze the performance metrics specific to each class to understand the effectiveness of the ML models used in this study. Specifically, for imbalanced datasets, the model's capacity to discriminate between classes may vary, substantially impacts performance metrics. To evaluate the model's overall efficacy and performance for each class, accuracy, recall, precision, and F1 scores were carefully examined. The following graphs were created to enable more accurate comparison and visual depiction of the model's performance by class. By highlighting the model's advantages and disadvantages for various classes, these visualizations provide insights for improvement strategies. As seen in Figure 4, the overestimation of the low class and the underestimation of the other classes are due to the imbalance of the class distribution in the dataset and how the model adapts to this imbalance. This is critical in understanding the model's learning bias towards certain classes, especially when working with imbalanced datasets. An imbalanced class distribution can cause the model to give more weight to the majority class and fail to learn rare classes well enough. Therefore, when evaluating the model's success for each class, it is important to examine how the predictions are distributed on a class basis. Different balancing techniques should be applied in line with these results, and performance improvement strategies should be developed to increase the model's sensitivity to imbalanced datasets.

Discussions

This study introduces a ML approach to predict the risk of DS using first-trimester combined screening test (FTS) data. The dataset analysis involves comparing various ML models, incorporating biochemical (hCG, PAPP-A) and biophysical (NT) parameters. AI has changed the world's agenda in recent years, and its use will become increasingly widespread in all areas in the coming years. There are few publications in the literature about the use of AI in obstetrics.

Neocleous et al.⁽²⁶⁾ developed an AI model that utilizes various features to assess the risk of aneuploidy and other chromosomal abnormalities. This model incorporates several variables, such



Figure 4. Performances of the classification models based on the proposed approach according to Down syndrome classes

as maternal age, the presence of a nasal bone, biochemical markers, (β -HCG, PAPP-A MoM), ultrasound measurements during pregnancy (CRL), NT, and a history of DS in prior pregnancies. These parameters were examined using ML algorithms to determine the probability of fetal abnormalities. The study emphasizes how important, ML is for identifying and evaluating the risks associated with chromosomal abnormalities. Koivu et al.⁽²⁷⁾ used the support vector machine (SVM) model to improve the precision of fetal DS screening. The SVM algorithm works exceptionally well with multidimensional data, making it ideal for intricate analyses such as identifying fetal abnormalities. Meanwhile, Subasi⁽²⁰⁾ conducted fetal

aneuploidy screening using non-invasive prenatal testing (NIPT) and fetal DNA (cffDNA) found in maternal blood. Their study significantly contributed to advancements in prenatal screening techniques by improving the accuracy of genetic tests and offering noninvasive alternatives to traditional diagnostic methods.

Durmuşoğlu et al.⁽¹¹⁾ proposed a ML model based on triple test indicators obtained from Gaziantep University, Turkey. This model aims to obtain more reliable results by avoiding the adverse effects of the triple test. The authors used nine ML models in the study and applied the SMOTE technique to generate synthetic data due to insufficient datasets. This

technique eliminated dataset imbalance and increased the model's accuracy.

Uzun and Kaya⁽²⁸⁾ developed an ML model, including Bayesian and naive Bayesian algorithms, using biochemical and biophysical FTS measurements to detect Trisomy 21. Since the dataset used in this study contained sample deficiencies and imbalances, the authors resorted to optimization techniques. This approach increased the model's performance and provided accurate results.

Catic et al.⁽²⁹⁾ proposed a neural network-based model for detecting DS and other genetic disorders (Edwards, Turner, Klinefelter Syndrome, Patau) using maternal serum screening data in the first trimester. They used a dataset of 2500 samples in their study, and the experimental results showed that recurrent neural networks provide higher accuracy than other methods. Wøjdemann et al.⁽³⁰⁾ used combined test data (NT and double test) and double test data (β -hCG, PAPP-A) to examine the detection of DS and other chromosomal abnormalities in a Danish population. The study aimed to enhance screening accuracy by evaluating the effectiveness of different test combinations. The findings highlight the significance of both dual and combined tests as essential tools for the early detection of chromosomal abnormalities.

These studies are important illustrations of the potential applications of technologies such as ML, AI, and DL in the healthcare industry, specifically in genetic disorder detection and prenatal screening. The high accuracy rates of the developed models can help reduce false positives and increase the dependability of screening results.

The study results show that machine learning-based models can be used effectively in prenatal screening processes and that higher accuracy rates can be achieved compared to traditional methods. Such AI-supported approaches can contribute to reducing unnecessary invasive tests and improving clinical decision-making processes by increasing early diagnosis accuracy. Future studies should aim to increase the generalization capacity of the model on larger data sets and to provide integration with different ML algorithms.

Study Limitations

This study has several limitations that should be acknowledged. First, although the dataset initially included 959 records, 106 cases (approximately 11%) were excluded due to missing or duplicate data. Even though data cleaning was carefully performed, this reduction may have caused a minor loss in statistical power. Internal comparisons between complete-case models and those with imputed data showed a slight average difference of 1.3% in accuracy. This suggests that missing data might have had a modest impact on model performance.

Second, due to the earthquake on February 6, 2023, our hospital was temporarily evacuated, and routine data archiving was disrupted. While test results before this date were systematically recorded in the WePoint system along with ultrasound images, the records after the earthquake lacked these images. As a result,

NT measurements in 318 cases could not be verified. Since NT is operator-dependent, this may have introduced measurement inconsistencies that could affect the model's accuracy.

Third, although the sample size was relatively large ($n=853$), all data were collected from a single tertiary hospital in southeastern Türkiye. Therefore, the findings may not fully represent populations from different geographic or clinical settings. Model performance may vary elsewhere, and recalibration could be needed before applying it in other regions.

Fourth, the study did not include diagnostic confirmation through invasive tests such as amniocentesis, nor did it include non-invasive tests like NIPT. Therefore, sensitivity, specificity, positive predictive value, and negative predictive value could not be calculated. In addition, follow-up data on post-screening clinical decisions were not available, limiting our ability to evaluate the real-world impact of the model.

Finally, this model was developed using only first-trimester biochemical and biophysical markers. Other clinical factors such as maternal health conditions, lifestyle habits, or family history were not included. Future research should include a wider range of clinical data and involve prospective, multicenter studies to improve generalizability and clinical usefulness.

Conclusion

Experimental findings indicate that tree-based ML models demonstrated superior performance, particularly in the presence of class imbalance within the dataset. Among the evaluated models, CatBoost achieved the highest accuracy rate at 95.31 percent, followed closely by XGBoost at 95.19 percent and LightGBM at 94.84 percent. These models were especially effective in capturing complex relationships among variables and showed strong generalization capabilities. In contrast, more traditional classification algorithms yielded comparatively lower accuracy scores, suggesting their limited capacity to handle the non-linear patterns and imbalance inherent in the data.

Further class-based performance analyses revealed that the dataset exhibited a skewed distribution, with a disproportionately high number of samples in the low-risk category. As a result, the models tended to overpredict the low-risk class while underrepresenting medium- and high-risk groups. This observation highlights a potential bias introduced by the class imbalance, which may compromise the model's sensitivity in detecting higher-risk cases.

To address this limitation, the integration of advanced data balancing techniques, such as oversampling, undersampling, or synthetic data generation, (e.g., SMOTE) is recommended for future research. Incorporating these methods may help enhance the model's performance across all risk groups, thereby improving both the fairness and diagnostic value of AI-supported prenatal risk assessments.

Ethics

Ethics Committee Approval: The study protocol and data collection process were reviewed and approved by the Çukurova

University Faculty of Medicine Research Ethics Committee in accordance with ethical standards (approval number: 144, date: 10.05.2024).

Informed Consent: Documents of informed consent were taken from all patients.

Footnotes

Authorship Contributions

Surgical and Medical Practices: E.Y., T.K.K., S.A., S.C.D., İ.C.E., M.S., M.A., F.İ.U., Concept: E.Y., T.K.K., S.A., S.C.D., İ.C.E., M.S., M.A., F.İ.U., Design: E.Y., T.K.K., S.A., S.C.D., İ.C.E., M.S., M.A., F.İ.U., Data Collection or Processing: E.Y., T.K.K., S.A., S.C.D., İ.C.E., M.S., M.A., F.İ.U., Analysis or Interpretation: E.Y., T.K.K., S.A., S.C.D., İ.C.E., M.S., M.A., F.İ.U., Literature Search: E.Y., T.K.K., S.A., S.C.D., İ.C.E., M.S., M.A., F.İ.U., Writing: E.Y., T.K.K., S.A., S.C.D., İ.C.E., M.S., M.A., F.İ.U.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

- Hodapp RM, Urbano RC, So SA. Using an epidemiological approach to examine outcomes affecting young children with Down syndrome and their families. *Downs Syndr Res Pract.* 2006;10:83-93.
- Rasmussen SA, Wong LY, Correa A, Gambrell D, Friedman JM. Survival in infants with Down syndrome, Metropolitan Atlanta, 1979-1998. *J Pediatr.* 2006;148:806-12.
- Cuckle H. Biochemical screening for Down syndrome. *Eur J Obstet Gynecol Reprod Biol.* 2000;92:97-101.
- Malone FD, Ball RH, Nyberg DA, Comstock CH, Saade GR, Berkowitz RL, et al. First-trimester septated cystic hygroma: prevalence, natural history, and pediatric outcome. *Obstet Gynecol.* 2005;106:288-94.
- Rifai N. Tietz Textbook of Clinical Chemistry and Molecular Diagnostics-E-Book: Tietz Textbook of Clinical Chemistry and Molecular Diagnostics-E-Book. Elsevier Health Sciences, 2017.
- Arnold KC, Flint CJ, Arnold KC, Flint CJ. Prenatal diagnostic testing for genetic disorders. *Obstetrics Essentials: A Question-Based Review.* 2017;237-43.
- Muñoz-Cortes M, Arigita M, Falguera G, Seres A, Guix D, Baldrich E, et al. Contingent screening for Down syndrome completed in the first trimester: a multicenter study. *Ultrasound Obstet Gynecol.* 2012;39:396-400.
- Allred SK, Takwoingi Y, Guo B, Pennant M, Deeks JJ, Neilson JP, et al. First trimester serum tests for Down's syndrome screening. *Cochrane Database Syst Rev.* 2015;2015:CD011975.
- Tamminga S, van Maarle M, Henneman L, Oudejans CB, Cornel MC, Siermans EA. Maternal plasma DNA and RNA sequencing for prenatal testing. *Adv Clin Chem.* 2016;74:63-102.
- Muller F, Thibaud D, Poloce F, Gelineau MC, Bernard M, Brochet C, et al. Risk of amniocentesis in women screened positive for Down syndrome with second trimester maternal serum markers. *Prenat Diagn.* 2002;22:1036-9.
- Durmuşoğlu A, Ay MM, Unutmaz Durmuşoğlu ZD. A classification model for predicting fetus with down syndrome—a study from turkey. *Applied Artificial Intelligence.* 2020;34:898-915.
- Aslan S. A deep learning-based sentiment analysis approach (MF-CNN-BILSTM) and topic modeling of tweets related to the Ukraine–Russia conflict. *Applied Soft Computing.* 2023;143:110404.
- Zhang H, Mo J, Jiang H, Li Z, Hu W, Zhang C, et al. Deep learning model for the automated detection and histopathological prediction of meningioma. *Neuroinformatics.* 2021;19:393-402.
- Cabello-Solorzano K, Ortigosa de Araujo I, Peña M, Correia L, J. Tallón-Ballesteros A. The impact of data normalization on the accuracy of machine learning algorithms: a comparative analysis. In *International Conference on Soft Computing Models in Industrial and Environmental Applications* (pp. 344-353). Cham: Springer Nature Switzerland, 2023.
- Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-94.
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Liu TY. Lightgbm: a highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30; 2017.
- Ibrahim AA, Ridwan RL, Muhammed MM, Abdulaziz RO, Saheed GA. Comparison of the CatBoost classifier with other machine learning methods. *IJACSA.* 2020;11.
- Zhang Z. Introduction to machine learning: k-nearest neighbors. *Ann Transl Med.* 2016;4:218.
- Webb GI, Keogh E, Miikkulainen R. Naïve Bayes. *Encyclopedia of Machine Learning.* 2010;15:713-4.
- Subasi A. Practical machine learning for data analysis using python. Academic Press, 2020.
- Cortes C. Support-vector networks. *Machine Learning*, 1995.
- Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. *JASTT.* 2021;2:20-8.
- Breiman L. Random forests. *Machine Learning.* 2001;45:5-32.
- Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences.* 1997;55:119-39.
- Aslan S. A novel TCNN-Bi-LSTM deep learning model for predicting sentiments of tweets about COVID-19 vaccines. *Concurr Comput.* 2022;34:e7387.
- Neocleous AC, Nicolaides KH, Schizas CN. First trimester noninvasive prenatal diagnosis: a computational intelligence approach. *IEEE J Biomed Health Inform.* 2016;20:1427-38.
- Koivu A, Korpimäki T, Kivelä P, Pahikkala T, Sairanen M. Evaluation of machine learning algorithms for improved risk assessment for Down's syndrome. *Comput Biol Med.* 2018;98:1-7.
- Uzun O, Kaya H. Prenatal risk assessment of Down's syndrome by probabilistic classifiers. In *Signal Processing and Communications Applications Conference (SIU)*, April 2013.
- Catic A, Gurbeta L, Kurtovic-Kozaric A, Mehmedbasic S, Badnjevic A. Application of neural networks for classification of Patau, Edwards, Down, Turner and Klinefelter syndrome based on first trimester maternal serum screening data, ultrasonographic findings and patient demographics. *BMC Med Genomics.* 2018;11:19.
- Wojdemann KR, Shalmi AC, Christiansen M, Larsen SO, Sundberg K, Brocks V, et al. Improved first-trimester Down syndrome screening performance by lowering the false-positive rate: a prospective study of 9941 low-risk women. *Ultrasound Obstet Gynecol.* 2005;25:227-33.