Risk Management and Healthcare Policy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Machine Learning-Based Prediction of First Trimester Down Syndrome Risk in East Asian Populations

Yen-Tin Chen[1,2], Gina Jinna Chen[3], Yu-Shiang Lin[1]

[1]In-Service Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; [2]Department of Obstetrics and Gynecology, Taipei Chang Gung Memorial Hospital, Taipei, Taiwan; [3]Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen, People's Republic of China

Correspondence: Yu-Shiang Lin, In-Service Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, No. 250, Wuxing St., Xinyi Dist., Taipei City, 110, Taiwan, Email eriklin@tmu.edu.tw

---

Purpose: Down syndrome is the most common chromosomal abnormality in newborns, often leading to developmental delays and congenital structural anomalies. This study employed multiple machine learning models to perform risk prediction and result exploration for first-trimester Down syndrome in East Asian populations, aiming to identify an optimal risk prediction model that will enhance future predictions of Down syndrome risk and improve the efficiency of the screening process.

Patients and Methods: This study collected data from the Down syndrome screening database at Taipei Chang Gung Memorial Hospital from May 1, 2018, to February 29, 2024. The dataset included 3,812 cases available for analysis, comprising 165 high-risk cases and 3,647 low-risk cases. Fourteen features (including maternal age, nuchal translucency thickness, serum markers, etc.) were input into the twelve machine learning models, along with seven data-balancing algorithms, to explore the risk prediction outcomes. The performance of these models was thoroughly evaluated using AUC (Area Under the Curve), accuracy, precision, recall, and F1 scores.

Results: Among the twelve machine learning models, the highest recall of 0.84 for high-risk cases was achieved by LightGBM combined with the RUS (Random Undersampling) data balancing algorithm. The highest AUC of 0.939 was attained by the ANN and LSTM models when combined with the ROS (Random Oversampling) data balancing algorithm.

Conclusion: The proposed ANN machine learning model, based on deep neural networks and combined with the ROS data balancing method, achieved an impressive AUC of 0.939 for classifying first-trimester Down syndrome risk in the East Asian population. Notably, this model also achieved an outstanding classification accuracy of 0.97. These results demonstrate the potential of the proposed ANN machine learning model for the accurate prediction of first-trimester Down syndrome risk.

Keywords: machine learning, first trimester down syndrome screening, deep neural network

---

Introduction

Down syndrome is a genetic disorder caused mainly by an extra chromosome 21. It is one of the most common chromosomal abnormalities in newborns. In addition to mild to moderate developmental delays,[1–4] individuals with Down syndrome may also exhibit specific congenital structural abnormalities, such as congenital heart disease (eg, ventricular septal defect),[5] gastrointestinal abnormalities,[6] and hypotonia (low muscle tone).[4] In Taiwan, the incidence of Down syndrome is approximately 7.92 per 10,000 live births,[7] meaning that about one in every 1,263 newborns is affected. Caring for individuals with Down syndrome requires significant effort and social resources, and it has a profound impact on their mothers and families.

The development of machine learning has made significant progress in recent years, particularly in its applications in medicine, such as image analysis,[8] cancer diagnosis,[9–11] diagnosis and progression of common diseases,[12–14] and cardiovascular disease risk prediction.[15,16] However, the application of machine learning in predicting the risk of Down syndrome is still limited. For the first-trimester Down syndrome screening, especially among East Asian

Received: 14 December 2024
Accepted: 23 March 2025
Published: 29 March 2025

populations, there is a lack of research exploring the use of machine learning models for risk prediction. Most studies have focused on using machine learning for second-trimester Down syndrome screening rather than first-trimester screening.

In a study on the second-trimester Down syndrome screening,[17] He et al utilized data from two hospitals in Beijing to train a machine-learning model for Down syndrome risk prediction. Data from Peking Union Medical College Hospital, which included 58,972 cases with 49 cases of Down syndrome, were used to train the model. External validation was conducted using data from Zibo Maternal and Child Health Hospital, which included 27,170 cases with 27 cases of Down syndrome. Second-trimester Down syndrome screening indicators, including AFP, uE3, free beta-HCG, maternal age, maternal weight, and gestational age, were used as features, with the first three indicators normalized to MoM (Multiples of the Median) values for training. The machine learning methods employed included Ensemble with Bootstrap Aggregating (bagging for short), Decision Tree, and Random Forest, with 10-fold cross-validation. The AUC for the internal validation set was 0.85, with a detection rate of 66.7% at a 5% false-positive rate and an increased detection rate of 75.0% when the false-positive rate was raised to 6%. For the results of the external validation set, the AUC was 0.89, with a detection rate of 85.2% at a 5% false-positive rate.

First-trimester Down syndrome screening predicts the risk of fetal Down syndrome by using maternal factors (such as age and ethnicity), fetal ultrasound (including nuchal translucency thickness and the presence or absence of the nasal bone), and maternal serum markers. Due to its high cost-effectiveness, first-trimester Down syndrome screening is considered highly important. However, it has timing constraints. The NT scan must be performed between 11 and 14 weeks of pregnancy; if this window is missed, the screening cannot be conducted. Previous studies have shown that the detection rate for the second-trimester Down syndrome screening is between 80–83%.[18,19] At the same time, Nicolaides et al have indicated that the detection rate for first-trimester screening can reach as high as 93–96%.[20,21] Machine learning (ML) models and traditional statistical approaches differ primarily in complexity, flexibility, and data requirements. Traditional statistical models, like logistic regression, are simpler, more interpretable, and rely on assumptions such as linearity and normality, making them suitable for smaller datasets and hypothesis testing. In contrast, ML models, such as random forests and neural networks, can capture complex, non-linear relationships without strict assumptions, often excelling with large, high-dimensional datasets. Thus, applying ML models to calculate feature values for the first-trimester Down syndrome screening and providing accurate predictions is more precise than using the second-trimester screening features.

Certain genetic factors may vary among different populations, so we limited our analysis to East Asians only. In the reference model proposed by the Fetal Medicine Foundation (FMF), East Asians were treated as a single group to eliminate inter-ethnic differences (the original model categorized individuals into White/Black/South Asian/East Asian groups) and restricting the study to East Asians further minimizes these differences. Additionally, there is a lack of research exploring the use of machine learning models for risk prediction in East Asians in the first trimester of Down syndrome screening.

Therefore, this study aims to achieve accurate risk prediction for the first-trimester Down syndrome screening by applying effective classification machine learning models. The primary advantage of machine learning is its ability to continually enhance classification capabilities by optimizing hyperparameters and model architecture. Additionally, feature selection algorithms can be employed to remove redundant features, improving the inference speed of the model and reducing the time obstetricians need to input clinical data. Furthermore, it can minimize the waste of resources in clinical testing, optimizing the overall screening process. In summary, utilizing machine learning for Down syndrome prediction offers significant clinical benefits, both in practical application and in the efficient use of resources.

This study utilized the first trimester Down syndrome screening database from Taipei Chang Gung Memorial Hospital from May 1, 2018, to February 29, 2024, to train twelve different machine learning models. The predictive outcomes provided by the software from the Fetal Medicine Foundation were used as the ground truth to evaluate the effectiveness of the twelve machine learning models proposed in this study. This research aims to establish an innovative approach based on machine learning to predict the first-trimester Down syndrome risk accurately.

The main contributions of this study are threefold: First, to our knowledge, this study represents the first study to use a dataset from a single East Asian population and apply machine learning models to investigate the first-trimester Down syndrome risk within this demographic, achieving the goal of precise diagnosis. Second, it proposes twelve different machine-learning models and conducts a detailed analysis of various classification metrics for each model. Third, due to the significant imbalance between the two populations (high risk/low risk) in this dataset, the study further explores the classification performance of these twelve machine learning models using seven different data balancing algorithms. It selects the most effective balancing algorithm and its corresponding classification model. Among the classification performances of the twelve machine learning models, the proposed ANN machine learning model based on deep neural networks, combined with the ROS (Random oversampling) data balancing method, achieved an ideal AUC (Area under curve) of 0.939 for the first-trimester Down syndrome risk classification in the East Asian population. Notably, this model also achieved an outstanding classification accuracy of 0.97. These results demonstrate the potential of the proposed ANN machine learning model for accurate first-trimester Down syndrome risk prediction.

Materials and Methods

Dataset

This study collected data from the Down syndrome screening database at Taipei Chang Gung Memorial Hospital from May 1, 2018, to February 29, 2024. Exclusion criteria included individuals of non-East Asian descent and cases with missing data. This study complied with the Helsinki Declaration and was approved by the Institutional Review Board of Chang Gung Medical Foundation on May 13, 2024, under protocol number 202400593B0. The informed consent was exempted by the Institutional Review Board of Chang Gung Medical Foundation, as it is a retrospective study utilizing a database, and the data has been anonymized, preventing any identification of individual cases.

Due to the low incidence of Down syndrome, we observed that high-risk cases are much fewer than low-risk cases in our database. This imbalance may adversely affect machine learning performance. Therefore, in addition to the original dataset, we employed seven data balancing algorithms to handle the disparity in case numbers between groups, including SMOTE (Synthetic Minority Over-sampling Technique) and its related algorithms (SVM-SMOTE, Borderline SMOTE, SMOTENC), ADASYN (Adaptive Synthetic Sampling), Random Under Sampling (RUS), and Random Over Sampling (ROS). We compared the performance of different machine learning models across various data-balancing algorithms and further compared them with the original dataset.

The study initially performed a random database split into training and testing datasets, with a ratio of 8:2. After separating the training and testing datasets, data balancing algorithms were conducted on the training dataset to prevent data leakage and ensure the independence of the test dataset. Finally, five-fold cross-validation was applied to ensure fairness and representativeness in model evaluation.

Data Preprocessing

We utilized binary classification to categorize Down syndrome risk into high-risk and low-risk groups. The predictive outcomes provided by the software from the Fetal Medicine Foundation were used as the ground truth. Cases with a Down syndrome risk of 1/270 or higher were labeled as 1 (high risk), while cases with a risk below 1/270 were labeled as 0 (low risk). The 1/270 risk threshold is primarily derived from historical data obtained from large prenatal screening studies, as well as guidelines from the American College of Obstetricians and Gynecologists (ACOG) and the UK National Screening Committee (UK NSC).[22] Continuous variables were standardized in the database using a Standard Scaler in Python. Categorical variables were converted into numerical form using One-Hot Encoding in Python to fit the format of machine learning algorithms. Missing values, or NaN (Not a Number), were handled based on the following two principles:

1. Features less relevant to Down syndrome: Maternal height and weight were imputed with the mean values of all cases. For maternal smoking and diabetes status, where over 99% of mothers in the database did not exhibit these factors, missing values were replaced with 0 (indicating non-smoking and no diabetes).
2. Features more relevant to Down syndrome: For features more associated with Down syndrome, including maternal age, fetal crown-rump length, fetal nuchal translucency (NT) thickness, and maternal serum markers (Free beta-hCG, PAPP-A), substitution with mean or other imputed values was not deemed appropriate. Therefore, cases with missing values for these features were excluded from the dataset.

Machine Learning Models

We conducted an in-depth analysis of the classification capabilities of twelve different machine learning models for first-trimester Down syndrome risk. In addition to traditional machine learning models, the study further incorporates three classification models based on deep neural networks—ANN, CNN, and LSTM—to examine whether deep neural network-based models can effectively improve the predictive accuracy of first-trimester Down syndrome risk. The twelve models used in this study are detailed below.

1. Logistic Regression
2. ElasticNet
3. RandomForest
4. AdaBoost
5. Gradient Boosting
6. eXtreme Gradient Boosting (XGBoost)
7. Light Gradient Boosting Machine (LightGBM)
8. Support Vector Machine (SVM)
9. Multilayer Perceptron (MLP)
10. Artificial Neural Network (ANN)
11. Convolutional neural network (CNN)
12. Long Short-Term Memory (LSTM)

Model Evaluation

Model performance was assessed using accuracy, the area under the receiver operating characteristic (ROC) curve (AUC), accuracy, precision, recall, and F1 score. All evaluation metrics were obtained from the test dataset. The evaluation indices are defined as follows:

$$Accuracy = \frac{TP + TN}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

$$P\mathrm{recision} = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$\mathrm{F1score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + Recall}$$

True positive (TP), true negative (TN), false positive (FP), and false negative (FN).

Since we focus on fetal Down syndrome risk prediction, the detection of high-risk groups is more critical, which means the primary goal of this study is to try to detect all fetuses with Down syndrome, minimizing false negatives. Therefore, in the effectiveness evaluation, emphasis will be placed on the high-risk group.

Platform and Statistics Analysis

The computer used for this study was equipped with an Intel 13600K processor, 32GB RAM, Nvidia 3090 24G GPU, and running on Windows 11. The analysis was conducted using Jupyter Notebook version 6.5.4. For statistical analysis, the scipy module in Python was used. Continuous variables were analyzed using the *t*-test and Mann–Whitney *U*-test, while categorical variables were analyzed using the chi-square test and Fisher's Exact Test.

Results

Baseline Characteristics

From May 2018 to February 2024, a total of 4,061 cases were collected. After excluding non-East Asian individuals (12 cases), 4,049 cases remained. The database includes age, height, weight, nulliparity, gestational age (days), fetal crown-rump length, nuchal translucency thickness, and serum data, as shown in Table 1.

Among all participants, the average age was 34.1 years, the average maternal height was 160 cm, and the average maternal weight was 57.7 kg. Of all the cases, 61.3% were first pregnancies. Only one individual was a smoker, and there were only 15 cases of pregestational diabetes. Most pregnancies were conceived naturally (83.1%), while 15.1% were conceived through in vitro fertilization (IVF). The average gestational age at the time of screening was 90.6 days, the average fetal length was 67.9 mm, and the average nuchal translucency thickness was 1.6 mm. The rate of nasal bone abnormalities was only 0.3%.

In the database, there were 4,049 East Asian individuals available for training. After preprocessing, cases with missing features that were more relevant to Down syndrome were excluded. Specifically, there were 63 missing entries for fetal nuchal translucency thickness, 232 missing entries for maternal serum free beta hCG, and 12 missing entries for maternal serum PAPP-A. After excluding these, 3,812 cases remained eligible for training, with 165 low-risk cases and 3,647 high-risk cases, as shown in Figure 1.

Model Setting

Twelve machine learning models were used to train and test their classification capabilities for first-trimester Down syndrome risk prediction. In addition to the nine traditional machine learning models, the study specifically designed

Table 1 Basic Characteristics of the Study Population

Characteristics	**All (*n*=4049)**
Age (y)	34.1 ± 2.8
Maternal Weight (kg)	57.7 ± 7.8
Maternal Height (cm)	160 ± 0.5
Fetal CRL (mm)	67.9 ± 4.1
Fetal NT (mm)	1.6 ± 0.2
Free beta hCG (IU/l)	75.6 ± 42.2
PAPP-A (IU/l)	4.6 ± 0.1
GA (days)	90.6 ± 2.1
Smoking	1 (0.02%)
Nulliparity	2482 (61.3%)
Pregestational Diabetes	15 (0.37%)
Conception	
Spontaneous	3364 (83.1%)
IVF	613 (15.1%)
Ovulation induction	72 (1.8%)
Nasal bone	
Normal	3175 (78.4%)
Abnormal	10 (0.3%)
Not Examined	864 (21.3%)

https://doi.org/10.2147/RMHP.S511035

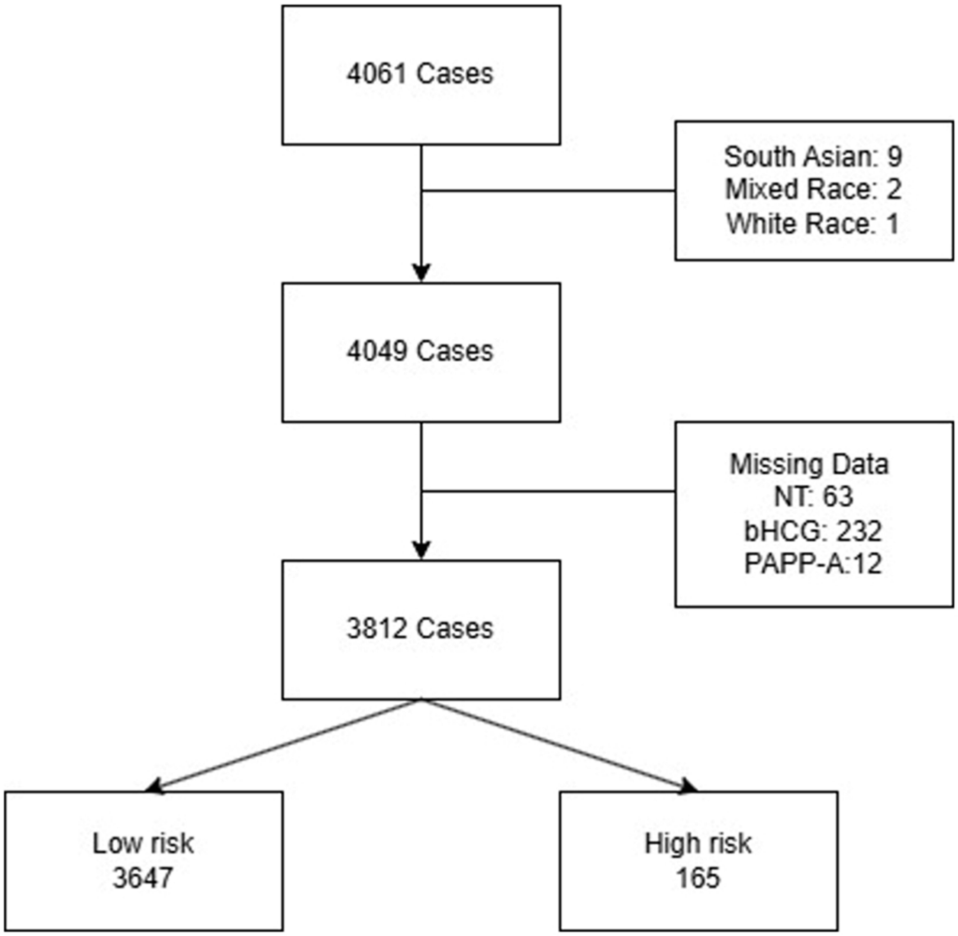

Figure 1 Flow diagram of patients included in this study.

three classification models based on deep neural networks to validate their ability to classify first-trimester Down syndrome risk. The three proposed deep neural network-based models are ANN, CNN, and LSTM, with detailed hyperparameter settings shown in Table 2.

During the training phase, the epoch for these three deep neural network-based models was set to 100, and an Early Stopping mechanism was implemented to prevent overfitting. Validation loss was used as the indicator, with the patience parameter set to 20 for ANN and 10 for CNN and LSTM.

Model Performance

Among all machine learning models, the best recall was 0.84, achieved by LightGBM combined with the random under-sampling data balancing algorithm. The best F1 score was 0.60, attained by LSTM combined with the SVM-SMOTE

Table 2 Hyperparameter of the Proposed Deep Neural Network Algorithms

Method	Hyperparameters
ANN	Layer=6, activation="ReLU and sigmoid", epoch=100, optimizer=Adam, learning_rate= 0.001, metrics=accuracy, epoch=100, Batch_size=128, Earlystopping(monitor=val_loss, patience=20)
CNN	Filter=32,32, kernel=3, activation="ReLU and sigmoid", epoch=100, optimizer= "Adam", learning_rate= 0.001, metrics=accuracy, epoch=100, Batch_size=128, Earlystopping(monitor=val_loss, patience=10)
LSTM	Units=50, time_step=1, learning_rate=0.001, activation=sigmoid, optimizer =Adam, loss_function=binary_crossentropy, metrics=accuracy, epoch=100, Batch_size=128, Earlystopping(monitor=val_loss, patience=10)

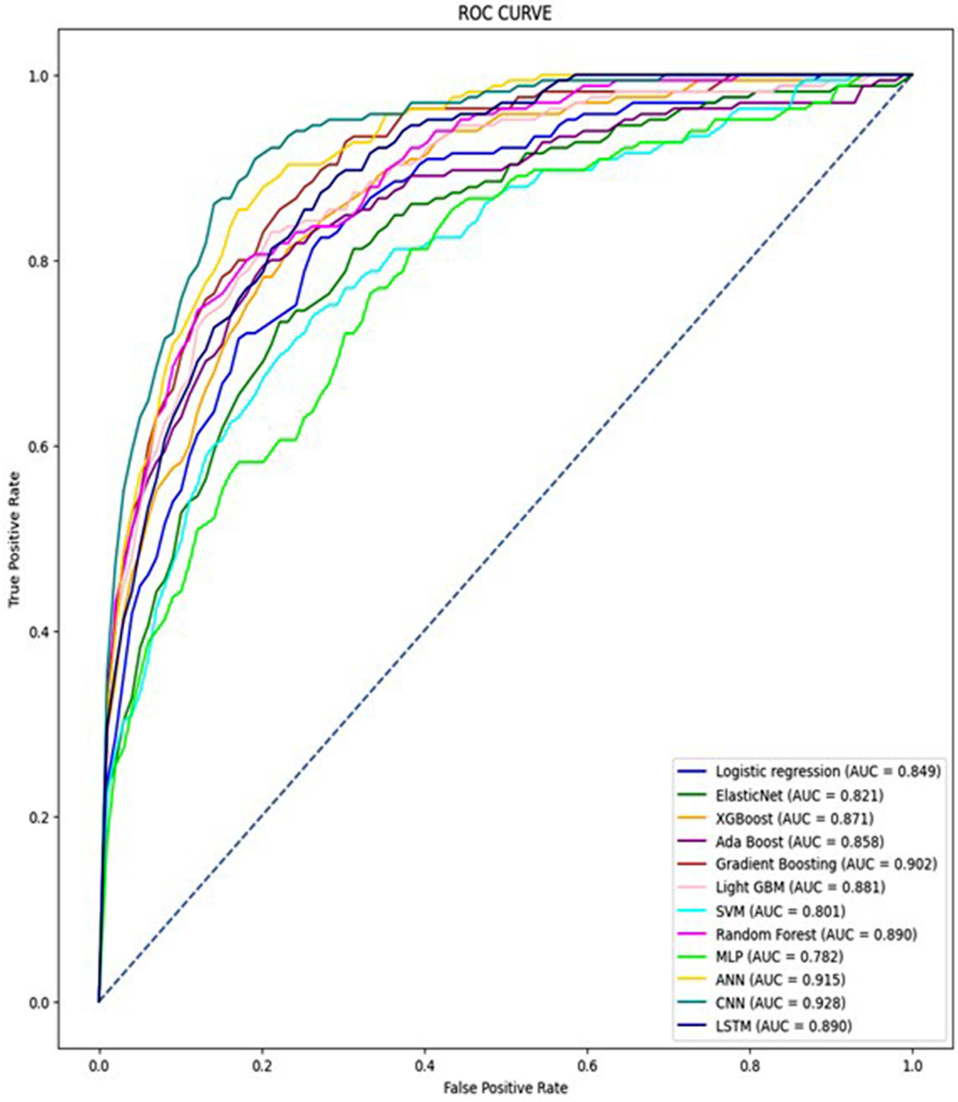

Figure 2 ROC Curves for twelve machine learning models on the original dataset.

data balancing algorithm. The highest AUC of 0.939 was achieved by ANN and LSTM combined with the ROS data balancing algorithm.

Figure 2 shows the ROC curve of the twelve machine learning models on the original dataset, where the best-performing model was CNN (AUC 0.928), followed by ANN (AUC 0.915) and Gradient Boosting (AUC 0.902). The worst-performing model was the MLP model (AUC 0.782). Figure 3 shows the AUC performance with the ROS data balancing algorithm applied. After applying ROS data balancing, most models exhibited improved AUC performance, with the best models being ANN and LSTM (AUC 0.939), followed by CNN (AUC 0.923). As the number of cases increases, the advantages of deep learning-based models in handling large datasets become more pronounced. As shown in Figure 3, the performance gap between the three deep learning-based models and conventional machine learning models becomes more significant. Additionally, the study found that after adjustment with ROS data balancing, the recall and F1 score for the high-risk group improved compared to the original dataset. Among all data-balancing algorithms, ROS had the best overall performance. For other algorithms, such as SMOTE and RUS, please refer to the appendices for further details (Tables S1-S8, Figures S1-S6). The settings for the other conventional machine learning models are shown in Table S9.

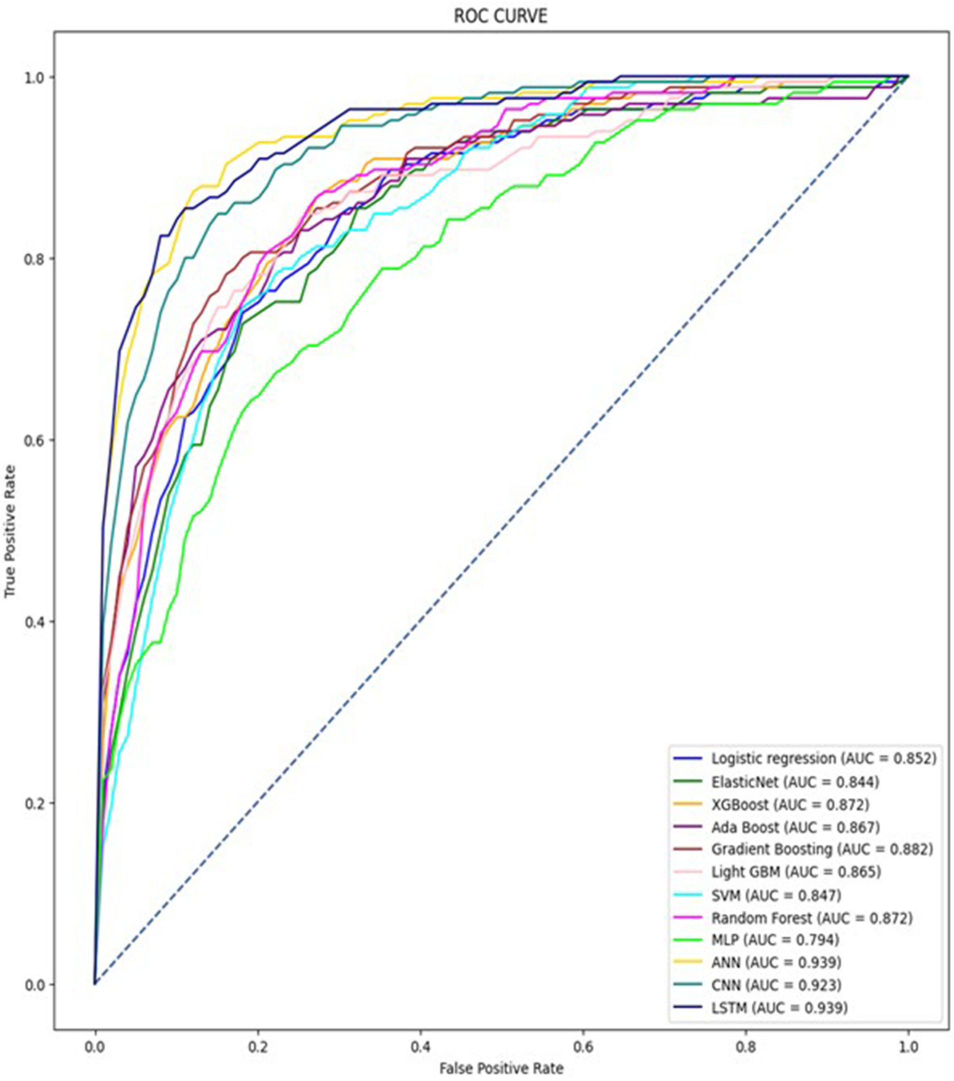

Figure 3 ROC Curves for twelve machine learning models on the random oversampling dataset.

As shown in Figure 4, the overall classification performance of machine learning algorithms combined with different data balancing algorithms was compared. Using AUC as the evaluation metric, we observed that the ROS data balancing algorithm combined with the CNN, ANN, and LSTM deep neural network architectures proposed in this study performed the best, achieving AUC values above 0.9. This indicates that combining deep neural networks with the ROS data balancing algorithm can effectively enhance classification performance. The original dataset's best-performing models were Gradient Boosting, ANN, and CNN. However, the Random Under-Sampling (RUS) data balancing algorithm demonstrated unsatisfactory performance when applied to CNN, ANN, and LSTM models. In addition, after applying imbalance handling algorithms such as SMOTE, ROS, and RUS, the results, as presented in Appendix Tables S1-S8, demonstrate a significant improvement in the sensitivity (recall) of most models for the high-risk group.

Discussion

This study applied the first-trimester Down syndrome screening database at Taipei Chang Gung Memorial Hospital. Twelve machine learning models and seven data balancing algorithms were used to explore the risk prediction outcomes. The best-performing models under five-fold cross-validation were the ANN and LSTM deep neural network models, combined with the random over-sampling data balancing algorithm, achieving an AUC of 0.939.

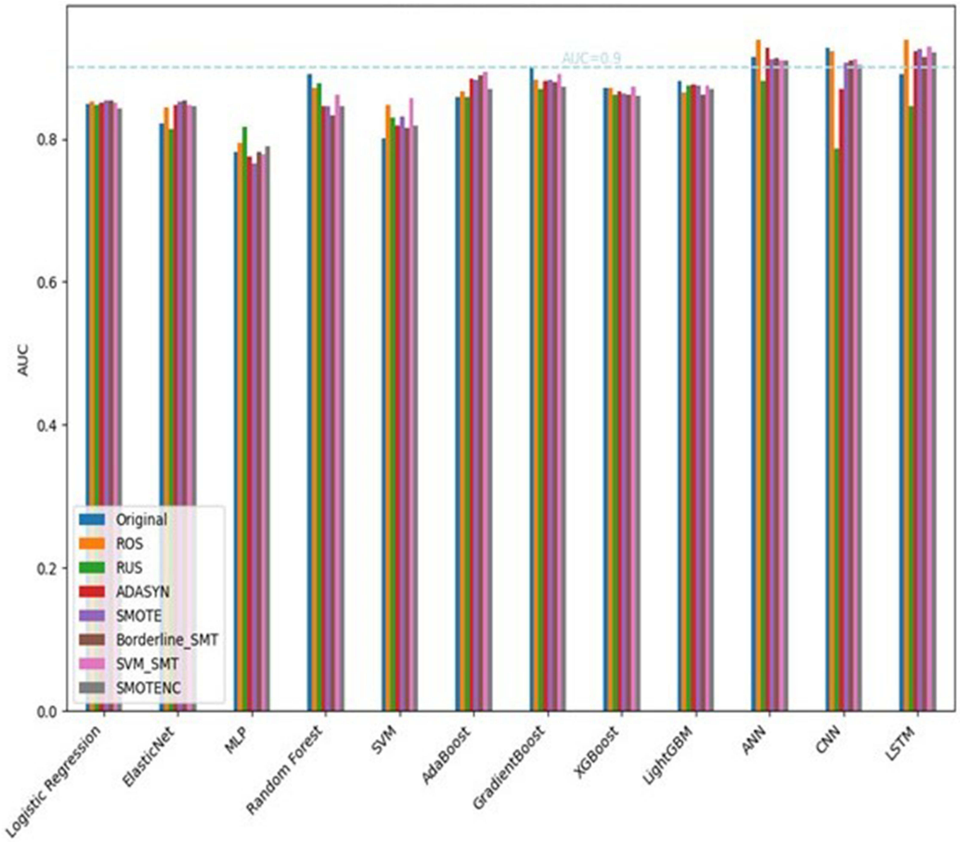

Figure 4 Comparison of AUC for different imbalance handling algorithms.

Since Down syndrome is not very common, high-risk cases are significantly fewer than low-risk cases, accounting for only 4.3% of the total in this database. As a result, machine learning models performed well on low-risk cases in the original dataset. Under five-fold cross-validation, the average precision across all machine learning models and datasets was 0.98, the recall was 0.99, and the F1-score was 0.98 for low-risk cases. However, performance on high-risk cases was subpar, with an average precision of 0.71, recall of 0.23, and F1-score of 0.33. Therefore, we applied several data balancing algorithms to the twelve machine learning models to improve performance for high-risk cases. We evaluated their classification performance and selected the most effective balancing algorithm and corresponding classification model.

Previous studies have found that the ANN model,[23–25] based on a deep neural network architecture, demonstrates high accuracy in predicting Down syndrome risk. The authors used ANN to achieve significantly higher AUC values. In particular, Neocleous et al showed that ANN could achieve a 100% detection rate for Down syndrome with a 4.8% false-positive rate.[25] Although these studies did not focus on East Asian populations, they highlighted the potential of ANN in Down syndrome risk prediction models.

Most studies used a database with confirmed Down syndrome cases and then compared their detection rate with the Fetal Medicine Foundation (FMF) model. In our approach, we used the FMF model as the ground truth and applied machine learning models to approximate its algorithm. Unlike the FMF model, which requires a fixed number of input features (more than 20 items needed to be input), our machine learning approaches, including the ANN-based model, offered the flexibility to reduce the number of input features. This adaptability enhances clinical workflow efficiency and reduces the burden of data collection.

In this study, we conducted an in-depth analysis of the East Asian population for the first time. Our findings showed that, in addition to the strong performance of ANN, another deep neural network model, LSTM, combined with the random over-sampling balancing method, achieved an equally excellent AUC of 0.939. Although ANN outperformed LSTM in terms of accuracy (0.97 vs 0.96), LSTM surpassed ANN in recall (0.78 vs 0.70). These results collectively demonstrate the superior classification ability of machine learning models based on deep neural network architectures for predicting first-trimester Down syndrome risk.

Deep learning-based models (eg, ANN, CNN, LSTM) often outperform conventional machine learning models due to their ability to automatically learn complex patterns and relationships from data without manual feature engineering. Unlike traditional models that rely on hand-crafted features, deep learning-based models, particularly deep neural networks with multiple layers, can capture hierarchical representations of data. This hierarchical learning allows deep learning-based models to understand intricate and non-linear patterns that conventional models like decision trees, support vector machines, or logistic regression might miss.

The first-trimester Down syndrome screening model proposed by the Fetal Medicine Foundation (FMF) remains accurate to this day, achieving a detection rate of 93–96%. Following the same logic, using features such as maternal age, nuchal translucency thickness, and serum markers to predict Down syndrome risk, the ANN model can achieve an AUC of 0.939. Moreover, the use of machine learning offers the advantage of flexibly reducing the number of input features, which can effectively improve clinical efficiency.

The limitation of this study is that we only included individuals of East Asian ethnicity. As a result, the machine learning model cannot calculate the risk of Down syndrome for other ethnic groups, such as Black, White, or South Asian populations. Therefore, this model cannot be generalized for use worldwide. Another limitation is that we used a binary classification approach rather than a regression model. The FMF model presents Down syndrome risk as numerical values, which is more intuitive and clinically helpful compared to simple classification. Since this study used binary classification, the interpretation of results may provide less information than numerical risk values.

In our future work, we aim to expand the database to identify actual Down syndrome cases and incorporate other chromosomal abnormalities, such as Edwards syndrome and Patau syndrome. Additionally, we plan to extend the dataset by including data from other hospitals in Taiwan and across East Asian countries, enhancing the model's applicability to the entire East Asian population. Furthermore, the interpretability of machine learning models is crucial for their practical clinical application. Future research can further explore methods to improve the explainability of Down syndrome risk prediction models, ensuring their clinical reliability and usability. We also plan to apply similar deep neural network-based machine learning models to provide a more comprehensive risk prediction. Moreover, since risk values are more intuitive when represented as numerical values, we will investigate advanced regression models in future studies to align risk prediction more closely with current clinical practice.

Conclusions

This study utilized a dataset of pregnant women from a single East Asian population and, for the first time, conducted an in-depth exploration of first-trimester Down syndrome risk using machine learning models specifically tailored for this population to achieve accurate diagnoses. Twelve machine learning models were proposed, and a detailed analysis of various classification metrics for each model was performed. Given the significant imbalance in the dataset between the two categories (high risk/low risk), the study further applied seven different data balancing algorithms to explore the classification performance of these twelve models, ultimately selecting the most effective balancing algorithm and its corresponding classification model. Among the classification performances of the twelve machine learning models, the proposed ANN model, based on a deep neural network architecture and combined with the ROS data balancing method, achieved an ideal AUC of 0.939 for classifying first-trimester Down syndrome risk in the East Asian population. Notably, the model also achieved an outstanding classification accuracy of 0.97. These results demonstrate the excellent potential of the proposed deep neural network-based ANN model for predicting first-trimester Down syndrome risk. In future studies, we aim to expand the database and establish a linkage between confirmed Down syndrome cases and high-risk cases identified in this study to further validate and enhance the model's accuracy. Furthermore, given that the numerical representation of Down syndrome risk is more intuitive for clinical interpretation, we plan to implement regression models to generate risk predictions that more closely align with current clinical screening practices. Lastly, as the interpretability of machine learning models is essential for their practical application in clinical settings, future research will focus on a comprehensive investigation into the interpretability of models used for predicting Down syndrome risk.

Ethics Approval

This study complied with the Helsinki Declaration and was approved by the Institutional Review Board of Chang Gung Medical Foundation on May 13, 2024, under protocol number 202400593B0. The informed consent was exempted by the Institutional Review Board of Chang Gung Medical Foundation, as it is a retrospective study utilizing a database, and the data has been anonymized, preventing any identification of individual cases.

Funding

This research was funded by the National Science and Technology Council, Taiwan (grant number NSTC 113-2222-E-038-001-MY3) and the Taipei Medical University (grant number TMU111-AE1-B30).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Bull MJ. Committee on G. Health supervision for children with down syndrome. *Pediatrics*. 2011;128(2):393–406. doi:10.1542/peds.2011-1605
2. Roizen NJ, Patterson D. Down's syndrome. *Lancet*. 2003;361(9365):1281–1289. doi:10.1016/S0140-6736(03)12987-X
3.. Brown FR 3rd, Greer MK, Aylward EH, Hunt HH. Intellectual and adaptive functioning in individuals with down syndrome in relation to age and environmental placement. *Pediatrics*. 1990;85(3 Pt 2):450–452.
4. Hall B. Mongolism in newborn infants. An examination of the criteria for recognition and some speculations on the pathogenic activity of the chromosomal abnormality. *Clin Pediatr*. 1966;5(1):4–12. doi:10.1177/000992286600500102
5. Irving CA, Chaudhari MP. Cardiovascular abnormalities in down's syndrome: spectrum, management and survival over 22 years. *Arch Dis Child*. 2012;97(4):326–330. doi:10.1136/adc.2010.210534
6. Fabia J, Drolette M. Malformations and leukemia in children with down's syndrome. *Pediatrics*. 1970;45(1):60–70. doi:10.1542/peds.45.1.60
7. Lin SY, Hsieh CJ, Chen YL. The impact of Down syndrome screening on Taiwanese down syndrome births: a nationwide retrospective study and a screening result from a single medical centre. *PLoS One*. 2013;8(9):e75428. doi:10.1371/journal.pone.0075428
8. You YS, Lin YS. A novel two-stage induced deep learning system for classifying similar drugs with diverse packaging. *Sensors*. 2023;23(16):7275. doi:10.3390/s23167275
9. Lin Y-S, Huang P-H, Chen -Y-Y. Deep learning-based hepatocellular carcinoma histopathology image classification: accuracy versus training dataset size. *IEEE Access*. 2021;9:33144–33157. doi:10.1109/access.2021.3060765
10. Mouelhi A, Rmili H, Ali JB, Sayadi M, Doghri R, Mrad K. Fast unsupervised nuclear segmentation and classification scheme for automatic allred cancer scoring in immunohistochemical breast tissue images. *Comput Methods Programs Biomed*. 2018;165:37–51. doi:10.1016/j.cmpb.2018.08.005
11. Yuan X, Xie L, Abouelenien M. A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. *Pattern Recogn*. 2017;77. doi:10.1016/j.patcog.2017.12.017
12. Wang L, Haug PJ, Del Fiol G. Using classification models for the generation of disease-specific medications from biomedical literature and clinical data repository. *J Biomed Informat*. 2017;69:259–266. doi:10.1016/j.jbi.2017.04.014
13. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*. 2017;5:8869–8879. doi:10.1109/ACCESS.2017.2694446
14. Buchan K, Filannino M, Ö U. Automatic prediction of coronary artery disease from clinical narratives. *J Biomed Informat*. 2017;72:23–32. doi:10.1016/j.jbi.2017.06.019
15. Lin C-M, Lin Y-S. TPTM-HANN-GA: a novel hyperparameter optimization framework integrating the Taguchi method, an artificial neural network, and a genetic algorithm for the precise prediction of cardiovascular disease risk. *Mathematics*. 2024;12(9). doi:10.3390/math12091303
16. Lin CM, Lin YS. Utilizing a two-stage Taguchi method and artificial neural network for the precise forecasting of cardiovascular disease risk. *Bioengineering*. 2023;10(11). doi:10.3390/bioengineering10111286
17. He F, Lin B, Mou K, Jin L, Liu J. A machine learning model for the prediction of down syndrome in second trimester antenatal screening. *Clin Chim Acta*. 2021;521:206–211. doi:10.1016/j.cca.2021.07.015
18. Schwartz S. Molecular cytogenetics and prenatal diagnosis. *Genetic Disorders Fetus*. 2015;313–347.
19. Cuckle H. Prenatal screening using maternal markers. *J Clin Med*. 2014;3(2):504–520. doi:10.3390/jcm3020504
20. Nicolaides KH. Screening for fetal aneuploidies at 11 to 13 weeks. *Prenat Diagn*. 2011;31(1):7–15. doi:10.1002/pd.2637
21. Nicolaides KH. Nuchal translucency and other first-trimester sonographic markers of chromosomal abnormalities. *Am J Obstet Gynecol*. 2004;191(1):45–67. doi:10.1016/j.ajog.2004.03.090
22. Wapner R, Thom E, Simpson JL, et al. First-trimester screening for trisomies 21 and 18. *N Engl J Med*. 2003;349(15):1405–1413. doi:10.1056/NEJMoa025273
23. Li L, Liu W, Zhang H, Jiang Y, Hu X, Liu R. Down syndrome prediction using a cascaded machine learning framework designed for imbalanced and feature-correlated data. *IEEE Access*. 2019;7:97582–97593. doi:10.1109/access.2019.2929681
24. Koivu A, Korpimaki T, Kivela 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. doi:10.1016/j.compbiomed.2018.05.004
25. Neocleous AC, Nicolaides KH, Schizas CN. Intelligent noninvasive diagnosis of aneuploidy: raw values and highly imbalanced dataset. *IEEE J Biomed Health Inform*. 2017;21(5):1271–1279. doi:10.1109/JBHI.2016.2608859