國立清華大學 電機工程學系

實作專題研究成果摘要

Detection of Increased Nuchal Translucency in First-Trimester Fetal Ultrasound Images Using a Two-Stage Deep Learning Model

以兩階段深度學習模型早期偵測胎兒超音波影像之頸部透明帶異常

專題領域： 系統領域

組　　別： A510

指導教授：陳國平

組員姓名：蔡秉言

研究期間：113年 7月 1日至 114年 5月 1日止，共10個月

Abstract

Increased nuchal translucency (NT) is a well-established marker for early screening of fetal chromosomal abnormalities such as Down syndrome. Nonetheless, conventional NT assessment necessitates the expertise of certified obstetricians, and the manual annotation involved is often time-consuming and susceptible to variability among different operators. To mitigate these issues, we collaborated with obstetricians at MacKay Memorial Hospital in Taipei to develop a deep learning-based assistive system that aims to provide high accuracy, objectivity, and interpretability.

In this study, we developed a two-stage deep learning framework. The first stage employs YOLO11n for fetal head localization and cropping, while the second stage utilizes DenseNet121 to classify whether the cropped image exhibits increased NT. The proposed model demonstrated excellent performance in both head detection and NT classification tasks. Notably, the classification model attained accuracy exceeding 99% on both the training and combined test datasets, with area under the receiver operating characteristic curve (AUC) values nearing 1.0.

In order to improve the transparency of the model, we employed Class Activation Mapping (CAM) techniques to visualize the model's attention. The resulting heatmaps consistently focused on the NT region, aligning with the diagnostic areas emphasized by clinicians, thereby enhancing the interpretability of the model.

These findings suggest that our two-stage framework holds strong potential for future integration into real-time clinical interfaces, serving as an effective and interpretable tool to assist obstetricians in early prenatal screening.

摘要

頸部透明帶(Nuchal Translucency, NT)增厚是早期篩檢胎兒染色體異常 (如唐氏症)的常用特徵。然而，傳統的 NT 量測需由具有認證資格的資深醫師操作，且後續的標註過程不僅耗時，也可能因操作者主觀差異而產生判讀不一致的情況。為了解決這些問題，我們與台北馬偕紀念醫院的婦產科醫師合作，嘗試利用深度學習建立一套具備高準確性、客觀與可解釋性的輔助判讀工具。

本研究實作了一個兩階段深度學習架構，第一階段使用 YOLO11n 模型自動偵測並裁切胎兒頭部影像，第二階段則以 DenseNet121 模型判斷裁切後的影像是否出現 NT 增厚的情形。實驗結果顯示，該系統在頭部定位與 NT 分類任務上皆展現優秀的表現，分類模型在訓練集與測試集上的準確率均超過99%，並在兩個資料集上皆達到接近1.0的 ROC 曲線下面積(AUC)分數。

此外，透過類別活化映射 (CAM)視覺化模型關注的區域，發現模型的注意力集中於 NT 位置，與臨床醫師關注的區域一致，證實了模型的可解釋性。

綜上所述，本研究所提出的兩階段模型具備準確、客觀與可解釋的特性，未來可望整合至臨床系統中，作為即時且有效的產前篩檢輔助工具。

1. Introduction

1-1. Background

Chromosomal abnormalities are a major cause of abnormal mental and physical development, with Down syndrome (trisomy 21) being the one of the most prevalent and recognized conditions [9]. Individuals diagnosed with Down syndrome frequently exhibit intellectual disabilities, and a variety of comorbid health issues requiring long-term care [1]. These conditions pose not only emotional and physical challenges for family caregivers, but also significant socioeconomic and healthcare burdens on society [3]. Therefore, the enhancement of screening tools employed in early pregnancy is essential.

Increased nuchal translucency (NT) is one of the most widely accepted non-invasive early screening markers for chromosomal abnormalities, detectable through ultrasound imaging [7]. NT refers to the subcutaneous fluid-filled space at the back of the fetal neck, and an increased thickness has been correlated with chromosomal abnormalities such as trisomy 21, 18, and 13 strongly [7]. However, measuring NT requires trained professionals, such as obstetricians with NT certification, to obtain reliable results. This measurement process is often time-consuming and may be influenced by operator-dependent variability, which can hinder its efficiency and consistency in routine screening practices.

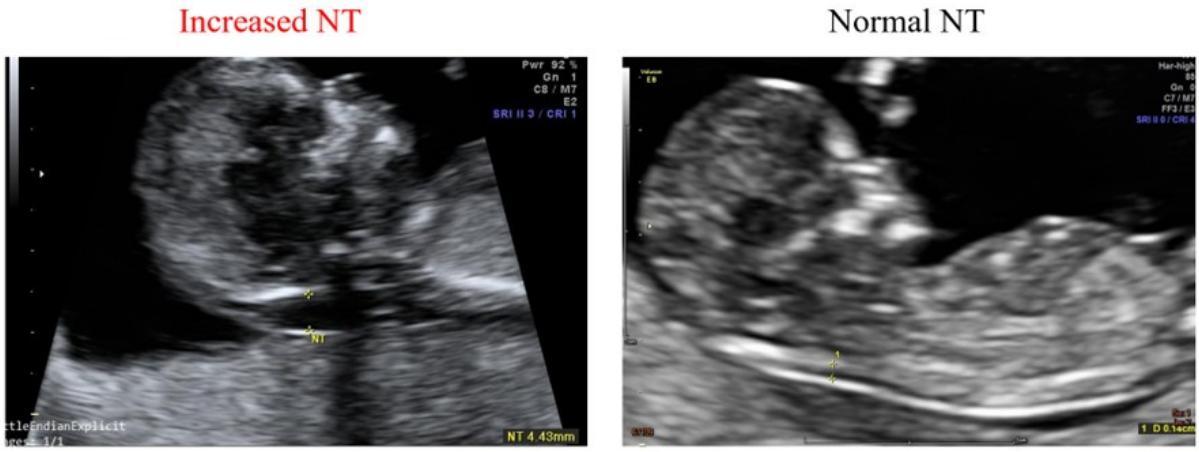

Fig. 1 Example of Increased and Normal Nuchal Translucency in First-Trimester Ultrasound Images

Convolutional Neural Networks (CNNs) have recently gained recognition as an effective tool for the analysis of medical ultrasound images, exhibiting notable efficacy in the identification of fetal markers. For example, Zhang et al. (2022) introduced a CNN-based screening model for the detection of trisomy 21, showing that deep learning is a feasible approach for prenatal screening using fetal ultrasound images [10]. Nonetheless, their approach

depended on the manual cropping of input images, a process that is both time-consuming and labor-intensive, and susceptible to inconsistencies among different operators.

1-2. Purpose

This study was conducted in collaboration with obstetricians at MacKay Memorial Hospital in Taipei, with the goal of addressing the challenges associated with traditional NT measurement. Our aim is to create a deep learning model that is characterized by high accuracy, objectivity, and interpretability. In the end, we aspire for this model to be incorporated into actual clinical processes as a useful aid for screening fetal chromosomal abnormalities during the first trimester

1-3. Method

In order to tackle these challenges, we referred to the research conducted by Tang et al. (2023), who introduced a novel three-stage model for the detection of prenatal genetic disorders [8]. Motivated by their methodology, we adopted a similar multi-stage design and developed a two-stage deep learning framework that first detects and crops the fetal head region using YOLO11n, subsequently utilizing DenseNet121 for the classification of increased NT.

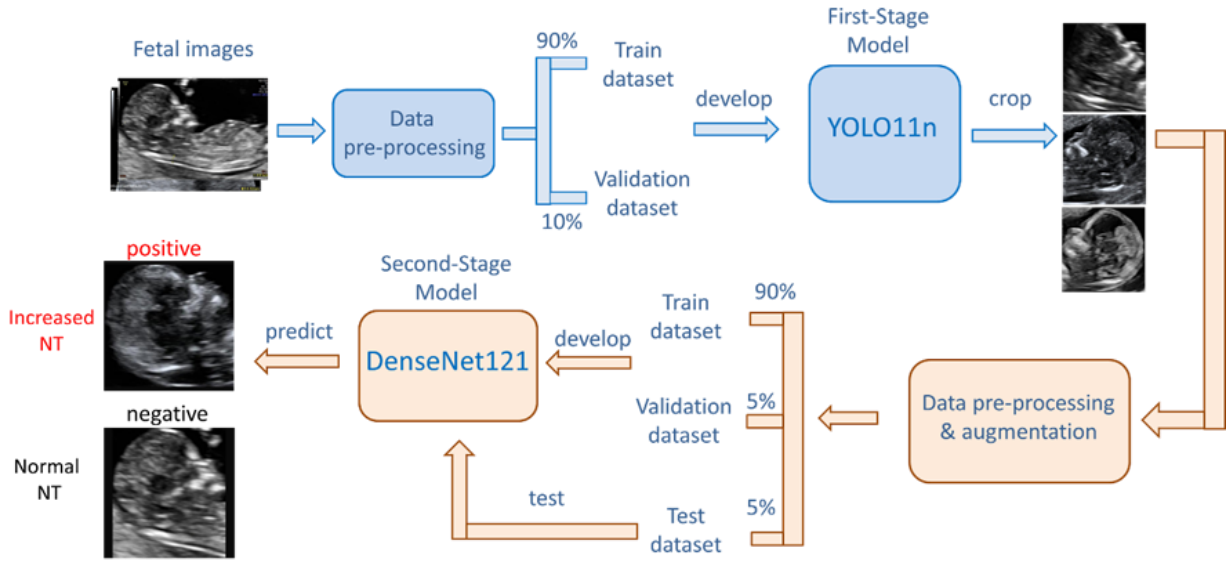

Fig. 2 Overview of the Proposed Two-stage Deep Learning Framework for Increased NT Detection.

As illustrated in Figure 2, the pipeline begins with pre-processing the raw ultrasound images. The processed images are first passed through a YOLO11n model to localize and crop the fetal head region. These cropped images then undergo further pre-processing and augmentation before being used to train a DenseNet121 classifier that determines whether NT is increased.

Additionally, to handle the issue of class imbalance, a prevalent issue in medical datasets where positive cases are frequently underrepresented, we utilized a weighted loss function during DenseNet121 training [11]. This approach enabled the model to pay more attention to clinically significant yet rare positive samples, thereby enhancing its sensitivity and robustness in practical screening applications.

1-3. Experimental Results

Figure 3 shows the confusion matrices for both the training set and the combined test set (validation + test set). The model correctly classified all cases in the combined test set, with only two misclassified negatives in the training set, resulting in nearly perfect classification performance.

In Figure 4, the Receiver Operating Characteristic (ROC) curves further validate the model's robustness, achieving an area under the curve (AUC) of 1.000 on both training and the combined test sets. This shows the model's excellent ability to distinguish between increased NT and normal cases.

Table 1 summarizes the performance of the fine-tuned DenseNet121 model on both the training set and the combined test set.

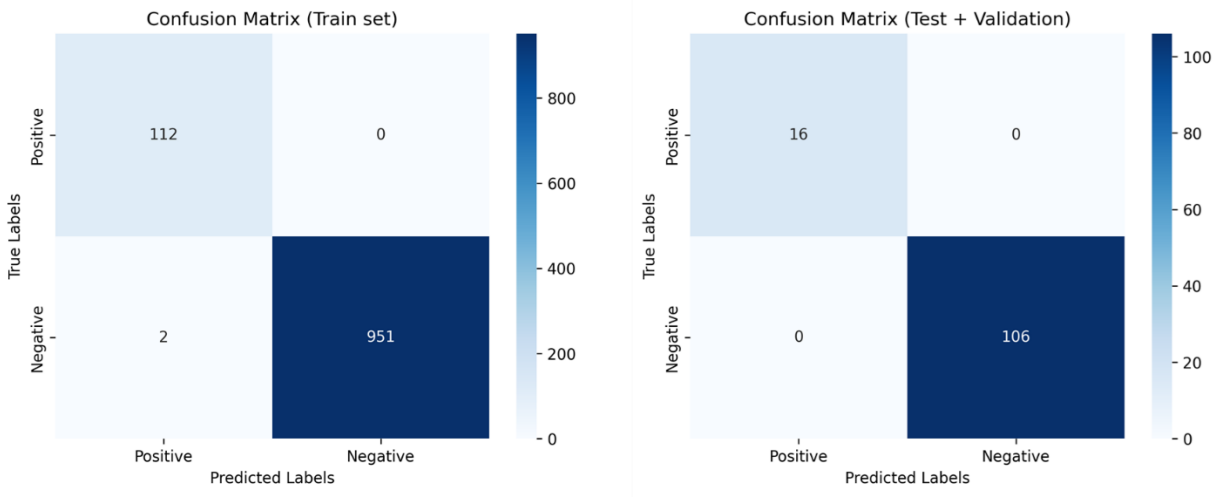

Fig. 3 Confusion Matrices on the Training set and Combined Test set

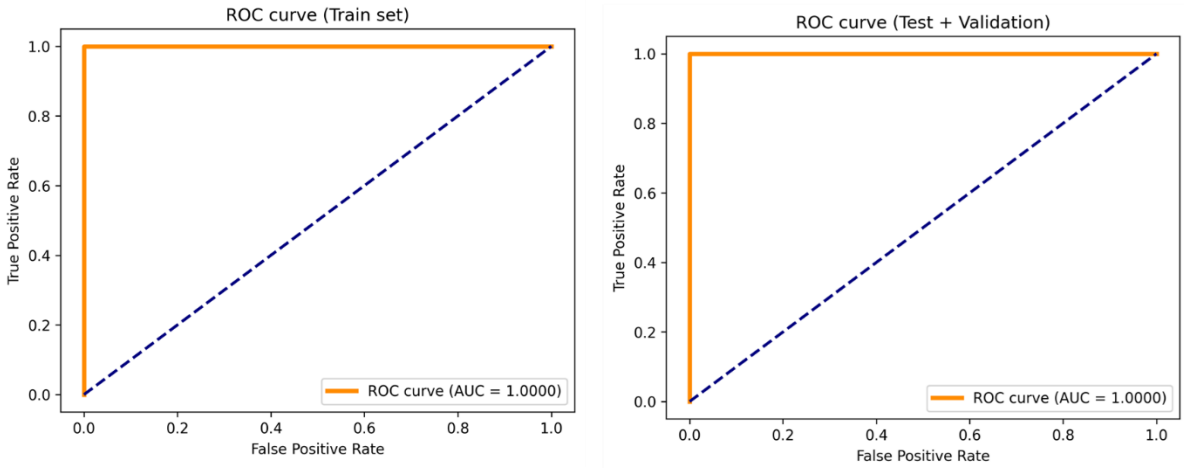

Fig. 4 ROC Curves of DenseNet121 on Training and Combined Test sets

Table 1

Performance Summary of the fine-tuned DenseNet121

Data	Type	ACC	AUC	Precision	Recall	F1_Score
Train Set	**Normal**	99.8%	100%	1.00	1.00	1.00
	Positive			0.98	1.00	0.99
	Total					1.00
Combined Test Set	**Normal**	100%	100%	1.00	1.00	1.00
	Positive			1.00	1.00	1.00
	Total					1.00

To evaluate the interpretability of our DenseNet121 model, we applied Smooth Grad-CAM++ on images randomly sampled from the combined test set. As illustrated in the examples of positive cases (Fig. 5, top row), the model consistently highlighted regions corresponding to the fetal nuchal translucency (NT), which aligns with the areas obstetricians typically focus on when assessing chromosomal abnormalities. Similarly, for negative cases (Fig. 5, bottom row), the model's attention also remained around the nuchal region.

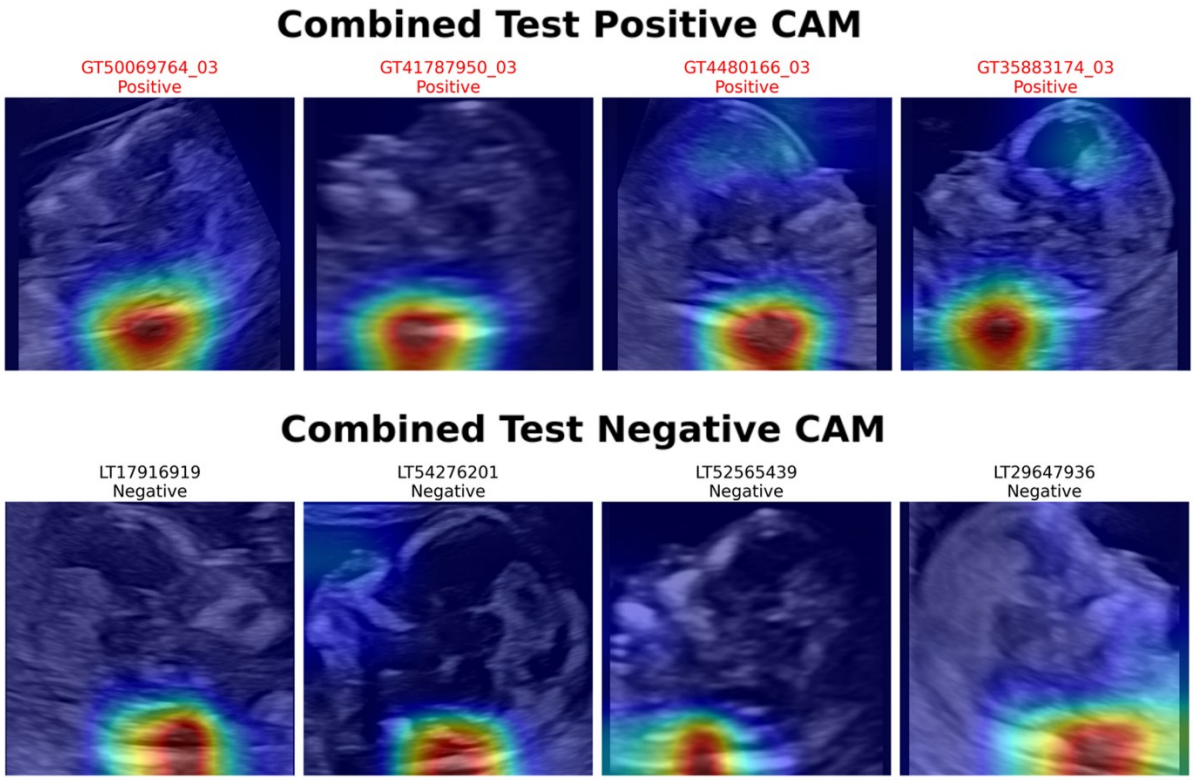

Fig. 5 Examples of Smooth Grad-CAM++ visualizations for randomly selected images of the combined test set. Top row: CAMs for ground true positive cases (increased NT); Bottom row: CAMs for ground true negative cases (normal NT).

1-5. Conclusion

In this project, we successfully developed a two-stage deep learning model to assist in the early screening of increased NT from first-trimester fetal ultrasound images. By integrating the lightweight YOLO11n for fetal head localization and DenseNet121 for NT classification, our framework effectively mirrors the clinical workflow while maintaining high performance and interpretability. The model demonstrated excellent accuracy, precision, recall, and AUC on both the training and the combined test datasets. Importantly, CAM visualizations showed that the model consistently focus on clinically relevant NT regions, further validating its reliability.

Looking ahead, this system holds strong potential for real-time clinical integration. However, our current software-only implementation on CPU suffers from noticeable delays when processing video inputs. Future work will therefore explore more lightweight yet accurate network architectures and consider hardware-level acceleration strategies such as deploying the model onto FPGA platforms to significantly reduce inference latency. These improvements will help bridge the gap between research and practical deployment into clinical screening support.

2. Reference

[1] **Asim, A., et al.** (2015). "Down syndrome: an insight of the disease." Journal of biomedical science 22: 1-9.

[2] **Chen, L., et al.** (2022). "Global, regional, and national burden and trends of Down syndrome from 1990 to 2019." Frontiers in Genetics 13: 908482.

[3] **Huang, G., et al.** (2017). Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition.

[4] **Khanam, R. and M. Hussain** (2024). "Yolov11: An overview of the key architectural enhancements." arXiv preprint arXiv:2410.17725.

[5] **Nicolaides, K. H.** (2004). "Nuchal translucency and other first-trimester sonographic markers of chromosomal abnormalities." American journal of obstetrics and gynecology 191(1): 45-67.

[6] **Omeiza,** D., et al. (2019). "Smooth grad-cam++: An enhanced inference level visualization technique for deep convolutional neural network models." arXiv preprint arXiv:1908.01224.

[7] **Taipale, P., et al.** (1997). "Increased nuchal translucency as a marker for fetal chromosomal defects." New England Journal of Medicine 337(23): 1654-1658.

[8] **Tang, J., et al.** (2023). "An Innovative Three-Stage Model for Prenatal Genetic Disorder Detection Based on Region-of-Interest in Fetal Ultrasound." Bioengineering 10(7): 873.

[9] **Theisen, A. and L. G. Shaffer** (2010). "Disorders caused by chromosome abnormalities." The application of clinical genetics: 159-174.

[10] **Zhang, L., et al.** (2022). "Development and validation of a deep learning model to screen for trisomy 21 during the first trimester from nuchal ultrasonographic images." JAMA network open 5(6): e2217854-e2217854.

[11] **Zergtant,** "Use weighted loss function to solve imbalanced data classification problems," Medium. [Online]. Available: https://medium.com/@zergtant/use-weighted-loss-function-to-solve-imbalanced-data-classification-problems-749237f38b75. [Accessed: May 1, 2025].

3. Reflections

This project was my first time independently executing a comprehensive deep learning pipeline from inception to completion. I did not expect the data preprocessing and YOLO annotation stages to be so time-consuming. Nonetheless, witnessing the final model achieve high performance proved to be immensely gratifying.

What brought me the greatest sense of fulfillment, however, was not just the performance metrics, but the process of discovering and solving key problems. A particularly noteworthy experience involved the realization that not all confirmed cases of Down syndrome exhibited clearly increased NT in the ultrasound images. In fact, some affected fetuses appeared almost indistinguishable from normal cases. Through repeated communication with our collaborating obstetrician, we clarified that these cases had not been identified during imaging but were subsequently diagnosed through more invasive screening techniques.

Upon reflection, I realized that this phenomenon is not unique to Down syndrome. In many diseases, it is common to see individuals with the same diagnosis exhibiting a wide spectrum of symptoms—some may be asymptomatic or have only mild signs, while others display more severe manifestations. Applying this perspective to fetal ultrasound imaging, it became clear that not all fetuses with Down syndrome would necessarily present with increased NT. I found this variation particularly fascinating, as it illustrates the complexity and heterogeneity of life, even in prenatal stages.

I would like to express my sincere gratitude to my advisor for offering such an interesting and meaningful research topic, and for consistently showing concern for our academic workload in other courses. I am also deeply thankful to the collaborating obstetricians for providing valuable data that made this project possible.

4. AI Usage Statement

Parts of this report were generated with the assistance of artificial intelligence (AI) tools. Specifically, AI was used to translate original Chinese sentences into English or to refine the wording of prewritten English sentences.