PJP Research Grant Report

Date of report: Dec 11 2024 ☐ This is a progress report ☒ This is a final report

Recipient Information

Investigator name and title	Dr. Brian Cox
Institution / Department	University of Toronto, Dept of Physiology
Address	1 Kings College St MS3360, Toronto Ontario, M5S 1A8
Phone number	416-978-3241
E-mail address	b.cox@utoronto.ca

Grant Details

Project name	Subtyping and clustering of preeclampsia, a personalized medicine approach to improve prediction			
Original grant amount & period	$ 90,500	From: 03/2022 (month & year)	To: 02/2024 (month & year)	
Has an extension been granted?	☒ Yes ☒ No	*If yes*, what is the current grant period?	From: 02/2024 (month & year)	To: 12/2024 (month & year)
Dates covered by this report	From: 12/2023 (month & year)	To: 12/2024 (month & year)		

Project Aims & Accomplishments

Short description of project & aims:

The objective of this project is to advance personalized medicine in preeclampsia, through the following aims:

1. Identify a consistent set of variables to subtype pregnancy for personalization of PE prediction
2. Build patient-specific, personalized corrections to biomarker levels
3. Identify risk factors of early-onset PE and early delivery using time-to-event modelling

Did you meet the aims you set out to achieve?

☐ Yes ☒ No

Please explain:

We made great strides pursuing Aim 1, capitalizing on an opportunity to make our analysis more comprehensive, to the full extent enabled by our novel patient subtyping methodology. Aim 2 was partly addressed in our work investigating the relationship between biomarkers and our subtypes. Moreover, we pivoted Aim 3 this project to a new aim following a new opportunity as we were able to recruit an expert specializing in Artificial Intelligence modelling who is interested in contributing to maternal and fetal health as a PhD trainee in our research group.

Have there been any significant changes to the original project plan over this reporting period? If so, how well have the project team managed these changes?

The team was very creative in looking for opportunities toi improve methods of analysis and accessing new sources of data. We secured a collaboration with an external research group that investigated best practices in patient subtyping and built an advanced software tool bespoke to our research question. Together we are co-authoring two publications on these methods that the Papas grant has contributed. With this direction, the depth of analysis has been widened and our means for validating generalizability has been improved, resulting in more rigorous analyses and more confidence in our findings. Our new expert Artificial Intelligence models used some of the funds from this grant to purchase a placenta image dataset to build computer vision artificial intelligence models that can accurately diagnose a preeclamptic pregnancy from a placenta image. We are hoping to eventually create a tool that can be used in remote and lower resource settings to support medical diagnosis in pregnancy.

Has the project faced any new ethical issues in the past year? If so, how well has the project addressed these?

No ethical issues

What did you accomplish during this reporting period?

We performed hypothesis-driven clustering to the SCOPE dataset to assess the clinical observability of the etiological subtypes identified in our previous molecular and histopathological work. Importantly this used only diagnostic criteria available ≤ 20 weeks of gestation. A clinically relevant subclassification tool will aid is sorting patients into different treatment groups. We observed a "canonical" cluster showing an alignment of low placental weight and a high uterine artery Resistance Index with more preterm births, more admissions to the neonatal unit, more small-for-gestational-age babies and the highest measurements of antepartum proteinuria; an "immunological" cluster showing an alignment of shorter length of sexual relationship with the father and admissions to the neonatal unit; and a "maternal" cluster showing an alignment of cardiovascular disease risk with higher Apgar Scores and fewer admissions to the neonatal unit (Figure 1, Table 2a and Table 2b). In addition, Plgf was found to be significantly lower in the canonical subtype. Finally, we developed risk groups from clinical and biomarker information at the 15-week and 20-week data collection timepoints which associate with significantly different disease trajectories. Overall, our results reinforce growing evidence that prognosis heterogeneity arises from distinct mechanisms, advancing patient subtyping as a promising avenue for developing better targeted diagnostic screens and treatment plans.

In Our new Aim we found that artificial neural networks, specifically vision transformers, are powerful enough to analyze placenta images captured using smartphone-grade cameras to detect pregnancy complications such as preeclampsia in low-resource settings, even in the absence of a perinatal pathologist. This suggests that in low-resource settings, where specialists and high-grade clinical devices are scarce, even non-clinical-grade devices can capture information on the pathology of the pregnancy and associated complications, when combined with artificial intelligence, can significantly advance maternal and obstetrical care.

For Final Report Only

What new information does this contribute to the field of preeclampsia?

Our work establishes the clinical utility of our etiological subtypes, organizing symptom heterogeneity into risk profiles that correlate with preeclampsia severity. Interestingly, a novel risk factor not previously identified in literature, a family-history of ischemic heart disease in the mother, appears to be a determining characteristic for severe maternal and fetal outcomes.

Finally, our computer vision models are among the first to demonstrate the potential of using a smartphone grade placental image and AI as a post-diagnostic tool, which can improve future health of mothers and neonates. Crucially, our tool can assist individuals in low-resource settings by determining which mothers and neonates need to be transported to high-resource settings for advanced care, based solely on images of the placenta.

Have you or do you plan to present this information at any meetings? Have you or will you submit a manuscript for publication?

Our findings on clustering subtypes of preeclampsia were been presented as a poster talk at the Southern Ontario Reproductive Biology Meeting (Toronto, Ontario, Canada). Further, the computer vision results were presented at the Computational and Artificial Intelligence Symposium at the International Federation of Placenta Association Conference (Montreal, Quebec, Canada).

We have also been invited to present our subtyping results to the 2025 Society for Reproductive Investigation Meeting (Charlotte, North Carolina, USA) as a poster talk.

Conversely, we co-authored two manuscripts sharing our new patient subtyping methodology to future biomedical researchers so they can study additional diseases; these papers are now in the submission process at *Nature Protocols* and the *Journal of Statistical Software*. We are currently drafting research articles to publish the full results of both analyses, to report the new insights on preeclampsia using our innovative methods.

What other impact has receiving this award had on you, your career, colleagues, and/or the field of preeclampsia?

The Papas award provided an opportunity to recruit new computation and bioinformatic support to our lab that would not have been possible otherwise. This increased research capacity of our group is helping move other preeclampsia related projects and provide new opportunities to apply for major research grants and awards. Our novel clustering work and AI driven vision models are being well received and may lead to new collaborations with researchers at other institutes. In particular recruiting an AI modelling expert is opening new research opportunities for work in other preeclampsia projects related to genetic.

Please provide a plain language summary of findings. This will be listed on the Preeclampsia Foundation website. *(350 words or less)*

Using the money from the Papas Award, we have addressed two critical issues in preeclampsia diagnosis and treatment research. Preeclampsia can be very different in each person affected; some become sick early in pregnancy, others later. Some have tiny growth-restricted babies, and some are normal-sized. Symptoms such as vision problems, abdominal pain, or trouble breathing can also be very different. We help create new statistical methods that can take complex data, such as patient histories, clinical data and blood work and identify what kind of preeclampsia a patient will develop. We name these preeclampsia types canonical, which starts early and with babies that are badly affected. Importantly, we found that this type is associated with a family history of heart disease. Another type we call immunological is when babies are affected, but mothers are not as strongly affected. This type is associated with first pregnancies and newer, shorter relationships. Lastly, we identified a type called maternal, where the babies appear unaffected. By predicting which kind of PE a person may develop, we can guide different treatment options and medical attention depending on the severity of the disease.

In another project, we are addressing how to diagnose disease in low-resource settings where medical expertise may not be easily accessible. Using artificial intelligence models, we created a tool that can predict preeclampsia from cell phone-quality pictures of a placenta after delivery. This can inform about risks for future pregnancies and whether the mother and child should be transported to a hospital for monitoring and medical assistance.

*Please email this completed Word document **and** a financial report to eleni.tsigas@preeclampsia.org*
If you have any questions, please email or call 321-421-6957

Supplementary Material

We repurposed patient information collected from the prospective study Screening for Pregnancy Endpoints (SCOPE) and filtered for samples which resulted in a preeclampsia diagnosis (n = 278). In addition to clinical associations found in previous histopathological and molecular studies, domain knowledge was used to prune variables according to relevance to the research question. The 32 in-model variables were grouped into 5 domains. Our strategy was to use our novel tool, similarity network fusion metaclustering, to perform the patient stratification. The premise behind similarity network fusion is to convert the variables into patient similarities which can each be represented by networks, then compressed into a single patient similarity network. The metaclustering facilitates the survey of the solution space from different sets of hyperparameters, allowing the identification of one or more well-represented "real" clustering solutions.

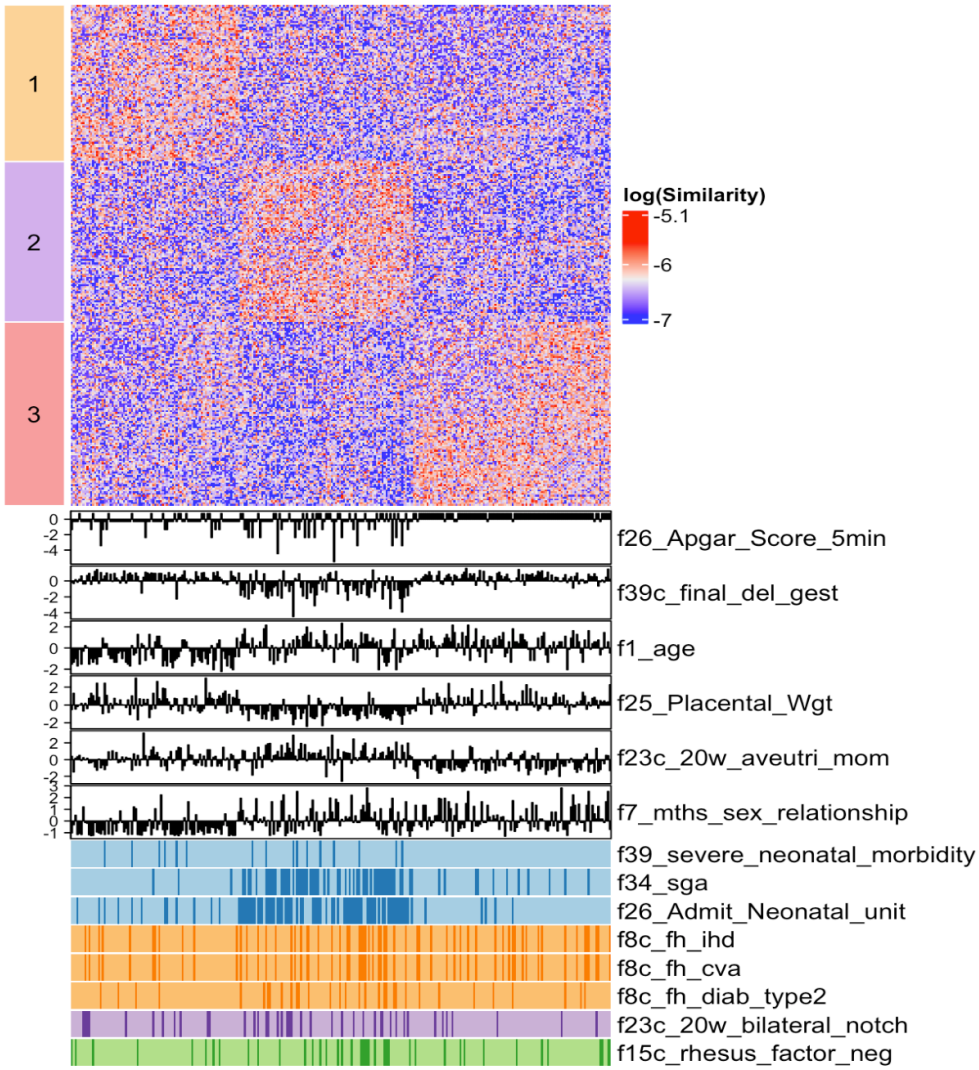

Figure 1. Patient-by-patient similarity heatmap of preeclamptic pregnancies, annotated with their clinical characteristics. The three clusters show distinct profiles which reflect the etiological subtypes found in our previous molecular and histopathological work. Cluster 1 appears to represent the immunological subtype, arising from immune-incompatibility with the fetus; Cluster 2 appears to represent the canonical subtype, arising from placental dysfunction; and Cluster 3 appears to align with the maternal subtype, arising from underlying cardiovascular disease.

There were 87 preeclamptics in Cluster 1, 89 in Cluster 2, and 102 preeclamptics in Cluster 3.

Table 1. Reference table for the variables included in the clustering analysis for investigating whether the three etiological subtypes are clinically observable.

Domain	Variable	Variable Description
Fetal Health	**f39_severe_neonatal_morbidity**	**Baby had severe neonatal morbidity or was a fetal death in utero or a stillbirth**
	f34_sga	**Small for gestational age (birthweight centile < 10th centile)**
	f26_Admit_Neonatal_unit	**Baby admitted to Neonatal unit**
	f26_Apgar_Score_1min	**Apgar score at 1 minute**
	f26_Apgar_Score_5min	**Apgar score at 5 minutes**
Cardiovascular Disease Risk	**f8c_fh_ch**	**Family history of chronic hypertension (participant's mother, father, sibling)**
	f8c_fh_vte	**Family history of venous thromboembolism (participant's mother, father, sibling)**
	f8c_fh_ihd	**Family history of ischaemic heart disease (participant's mother, father, sibling)**
	f8c_fh_cva	**Family history of cerebrovascular accident (participant's mother, father, sibling)**
	f8c_fh_diab_type1	**Family history of diabetes type 1 (participant's mother, father, sibling)**
	f8c_fh_diab_ type2	**Family history of diabetes type 2 (participant's mother, father, sibling)**
	f1_age	**Participant's age**
	f11c_waist_hip_ratio	**Waist-to-hip ratio at 15 weeks**
	f11_bmi	**BMI at 15 weeks**
Placenta	**f25_Placental_Wgt**	**Placental weight**
	f23c_20w_umbri_mom	**Umbilical artery Doppler Resistance Index at 19-21 weeks, transformed to MoM by gestation**
	f23_20w_umbpi	**Umbilical artery Doppler Pulsatility Index (PI) at 19-21 weeks**
	f23c_20w_aveutri_mom	**Uterine artery Resistance Index at 19-21 weeks**
	f23c_20w_bilateral_notch	**Bilateral notch at 19-21weeks**
	f23c_20w_unilateral_notch	**Unilateral notch at 19-21 weeks**
Maternal Health	**f24_proturia_dipstick**	**Highest pre-labour proteinuria measured by dipstick**
	f24_Urin_prot_creat_ratio_mmol	**Highest pre-labour urinary protein/creatinine ratio (mg/mmol) at the end of pregnancy**
	f39c_final_del_gest	**Gestational Age at Delivery (weeks)**
	bb_total_hdl_ratio	**Biobank Total cholesterol: HDL cholesterol ratio**
	ratio_sBP	**Ratio of systolic blood pressure (20-week to 15-week)**
	ratio_dBP	**Ratio of diastolic blood pressure (20-week to 15-week)**
Immune Activation	**f7_mths_sex_without_barrier**	**Months of sexual relationship without barrier contraception with biological father of baby**

f7_mths_sex_relationship	**Months of sexual relationship prior to conception with the biological father of the baby**
f7_sex_per_mth_b4_preg	**Frequency of sexual intercourse with biological father of baby per month in the 3 months prior to conception**
f9c_any_infection	**Any infection (urti or uti or pyelnephritis or gastro or vag candida or other infections) in pregnancy before 15 weeks**
f15c_rhesus_factor_neg	**Participant's rhesus factor**
f15c_vag_any_infect_gp	**Severity of vaginal infection on vaginal swab at <20 weeks**
f18c_any_infection	**Any infection (urti or uti or pyelnephritis or gastro or vag candida or other infections) between 15-weeks and 20-weeks**

Table 2a. Clinical profiles of each cluster. These numerical variables were analysed for significant differences in distribution between the clusters using a Kruskal-Wallis test and their Eta-squared effect size.

Domain	Variable	H Statistic	P-value	Adjusted P-value	Eta^2	CI low	CI high	Cluster 1	Cluster 2	Cluster 3
Fetal Health	f26_Apgar_Score_1min	47.824	0.000	0.000	0.167	0.104	1.000	8.149	7.157	8.755
	f26_Apgar_Score_5min	111.344	0.000	0.000	0.398	0.339	1.000	9.000	8.820	9.922
Cardiovascular Disease Risk	f1_age	52.449	0.000	0.000	0.183	0.115	1.000	24.011	28.831	29.784
	f11c_waist_hip_ratio	7.560	0.023	0.032	0.020	0.000	1.000	0.848	0.847	0.825
	f11_bmi	4.083	0.130	0.156	0.008	0.000	1.000	29.194	27.504	26.778
Placenta	f25_Placental_Wgt	95.229	0.000	0.000	0.339	0.260	1.000	637.195	424.899	637.510
	f23c_20w_umbri_mom	2.260	0.323	0.334	0.001	0.000	1.000	1.016	1.017	0.990
	f23_20w_umbpi	2.192	0.334	0.334	0.001	0.000	1.000	1.311	1.330	1.273
	f23c_20w_aveutri_mom	63.081	0.000	0.000	0.222	0.143	1.000	1.073	1.193	0.954
Maternal Health	f24_proturia_dipstick	6.713	0.035	0.045	0.017	0.000	1.000	2.437	2.876	2.863
	f24_Urin_prot_creat_ratio_mmol	8.696	0.013	0.021	0.024	0.002	1.000	97.609	213.562	122.284
	f39c_final_del_gest	74.610	0.000	0.000	0.264	0.182	1.000	39.103	35.766	39.080
	bb_total_hdl_ratio	3.336	0.189	0.212	0.005	0.000	1.000	3.331	3.456	3.160

	ratio_sBP	11.403	0.003	0.006	0.034	0.004	1.000	1.033	1.040	0.995
	ratio_dBP	8.403	0.015	0.022	0.023	0.002	1.000	1.024	1.028	0.983
Immune Activation	f7_mths_sex_without_barrier	29.921	0.000	0.000	0.102	0.059	1.000	22.655	44.551	49.386
	f7_mths_sex_relationship	47.765	0.000	0.000	0.166	0.110	1.000	30.379	61.444	64.934
	f7_sex_per_mth_b4_preg	47.089	0.000	0.000	0.164	0.097	1.000	23.621	11.397	9.921

Table 2b. Clinical profiles of each cluster. These binary categorical variables were analysed for significant differences in distribution between the clusters using a Chi-square test and their Cramer's V effect size.

Domain	Variable	Test Statistic	P-Value	Adjusted P-value	Cramer's V	CI Low	CI High	Cases	Cluster 1	Cluster 2	Cluster 3
Fetal Health	f39_severe_neonatal_morbidity	11.366	0.003	0.010	0.184	0.030	1.000	16	7%	11%	0%
	f34_sga	100.319	0.000	0.002	0.596	0.492	1.000	70	3%	63%	11%
	f26_Admit_Neonatal_unit	118.798	0.000	0.002	0.649	0.546	1.000	83	15%	73%	5%
Cardiovascular Disease Risk	f8c_fh_ch	6.633	0.038	0.058	0.129	0.000	1.000	132	38%	57%	47%
	f8c_fh_vte	2.534	0.272	0.314	0.044	0.000	1.000	29	9%	15%	8%
	f8c_fh_ihd	8.172	0.018	0.046	0.149	0.000	1.000	63	13%	30%	25%
	f8c_fh_cva	6.859	0.035	0.058	0.132	0.000	1.000	24	2%	10%	13%
	f8c_fh_diab_type1	0.183	1.000	1.000	0.000	0.000	1.000	8	2%	3%	3%
	f8c_fh_diab_type2	15.927	0.000	0.002	0.224	0.099	1.000	32	5%	22%	8%

Placenta	f23c_20w_bilateral_notch	14.064	0.001	0.006	0.209	0.078	1.000	41	13%	26%	7%
	f23c_20w_unilateral_notch	1.953	0.392	0.420	0.000	0.000	1.000	49	18%	21%	14%
Immune Activation	f9c_any_infection	6.520	0.037	0.058	0.128	0.000	1.000	107	49%	35%	32%
	f15c_rhesus_factor_neg	7.196	0.022	0.048	0.137	0.000	1.000	40	9%	22%	12%
	f15c_vag_any_infect_gp	4.558	0.083	0.114	0.096	0.000	1.000	6	2%	4%	0%
	f18c_any_infection	3.537	0.168	0.210	0.074	0.000	1.000	75	32%	29%	21%

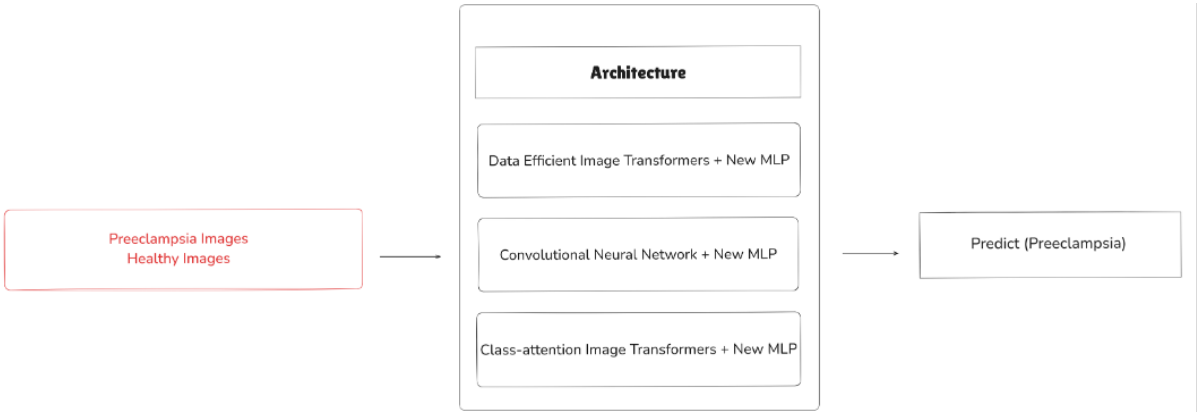

Figure 2. Graphical summary of the neural network architecture used to build the computer vision models that can identify whether a placenta is associated with a preeclamptic versus health pregnancy.

Our team developed macroscopic placenta image classifiers using neural networks to detect preeclampsia in low-resource settings. The motivation behind this project was the limited availability of perinatal pathologists in low-resource settings to analyze or assess the placenta, a procedure crucial for maternal-fetal health during pregnancy and after delivery. Therefore, we investigated the use of artificial neural networks as an image analysis tool to develop models that address this gap. Additionally, it was important to do this as a step towards addressing preeclampsia because the disorder disproportionately affects individuals in low-resource and income settings.

During this reporting period, we implemented and trained three maternal-side placental models and three fetal-side placental models using three different backbone neural network architectures: data-efficient image transformers and class-attention image transformers. Crucially, to develop the maternal-side placental models, we used a total of 384 maternal placental images (192 preeclampsia and 192 non-preeclampsia images). Similarly, to develop the fetal-side placental models, we used a total of 384 fetal placental images (192 preeclampsia and 192 non-preeclampsia images). For both maternal-side and fetal-side placental models, a 90% training and 10% testing split was employed. Additionally, we preprocessed the images by resizing them, applying random rotations and flips, and normalizing them. Subsequently, we trained the models using 5-fold cross-validation on the training set and evaluated their performance on the test set.

Importantly, the performance of the classifiers was evaluated using accuracy, precision, recall, F1 score, and AUC scores. Our final findings suggest that preeclampsia can be accurately detected from both maternal-side and fetal-side placental images using neural networks. Furthermore, we identified that the current state-of-the-art neural networks, namely transformers, perform better at detecting preeclampsia than convolutional neural networks when given both maternal-side and fetal-side images. Interestingly, we discovered that data-efficient image transformers achieved the highest accuracy, precision, recall, F1 score, and AUC scores based on maternal-side placenta images. Similarly, class-attention image transformers demonstrated superior performance with fetal-side placental images. Overall, we found that there were fewer misclassifications with the

transformer models compared to the convolutional neural network models. Our results are summarized in Table 1 below.

Table 3. Mean performance metrics for all models

Model	Mean Accuracy	Mean Precision	Mean Recall	Mean F1 Score	Mean AUC-ROC Score
CNN(F)	0.8051	0.8167	0.8051	0.8037	0.8705
CaiT(F)	0.8821	0.8863	0.8821	0.8816	0.9695
DeiT(F)	0.8308	0.8414	0.8308	0.8273	0.9174
CNN(M)	0.759	0.7766	0.759	0.7546	0.8558
CaiT(M)	0.8051	0.8071	0.8051	0.8049	0.9016
DeiT(M)	0.9333	0.9377	0.9333	0.9332	0.9879

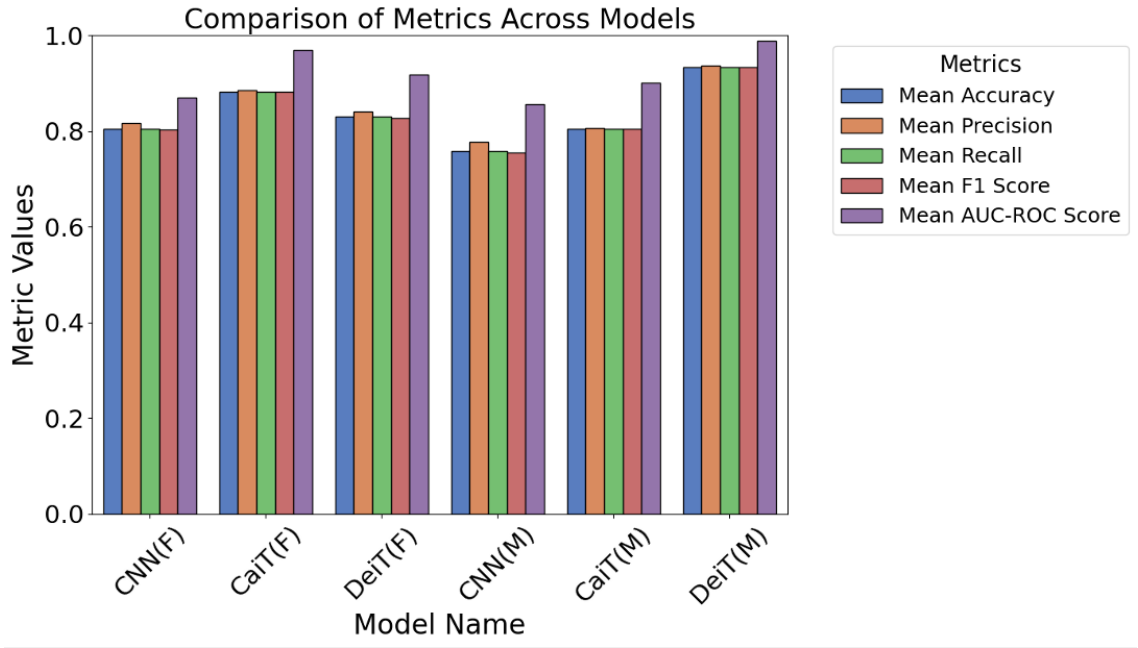

Figure 3. Barplot visualizing the difference in performance across all the models for accurately identifying a preeclamptic placenta.

Figure 4a. Comparison between maternal placental side and fetal placental side Data-efficient Image Transformer models

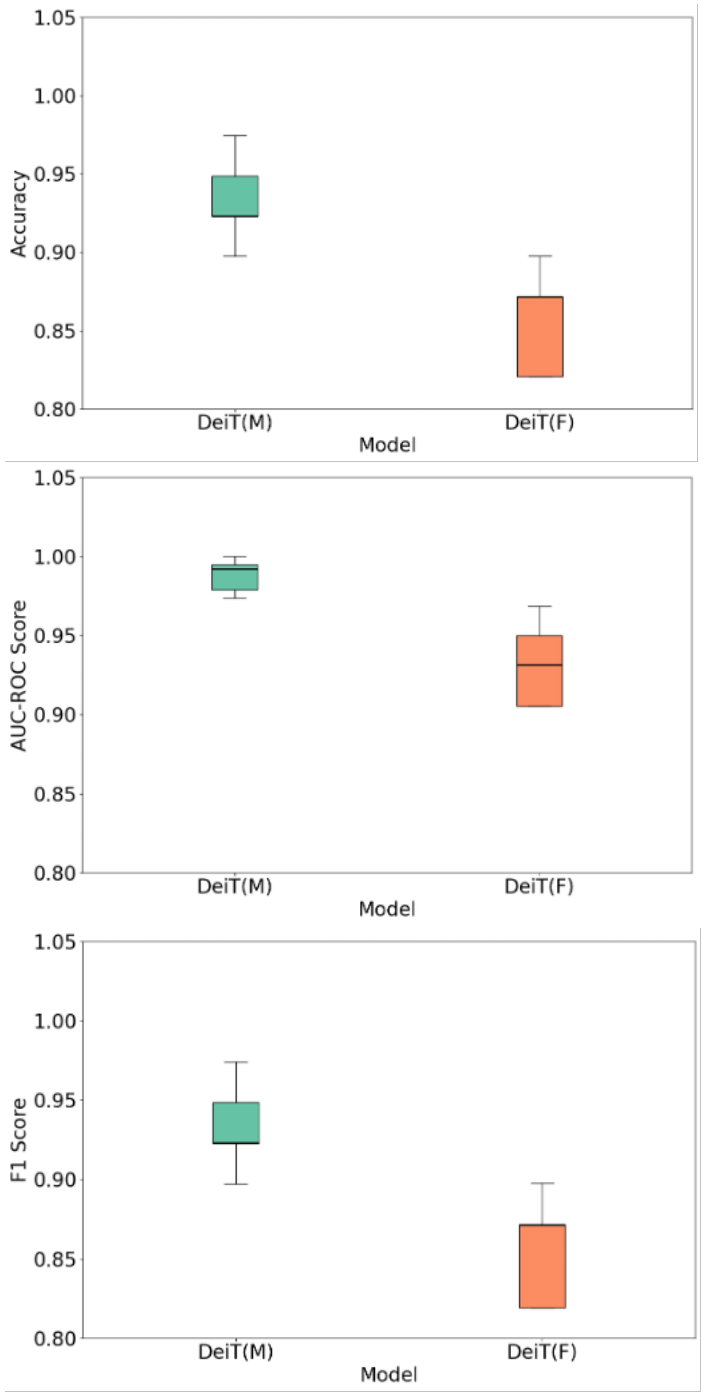

Figure 4b. Comparison between maternal placental side Convolutional Neural Network, Data-efficient Image Transformer, and Class-attention Image Transformer models

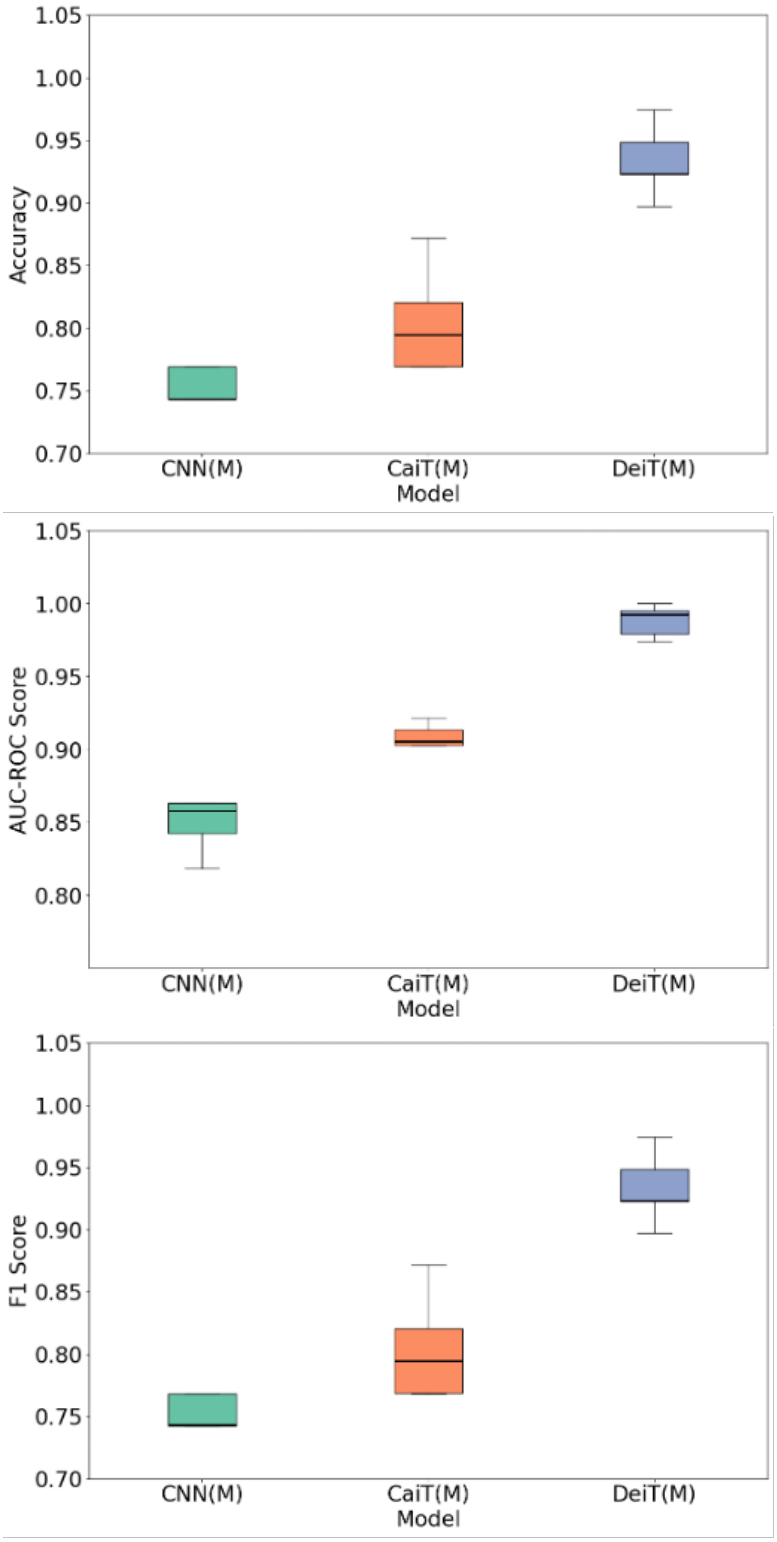

Figure 4c. Comparison between fetal placental side Convolutional Neural Network, Data-efficient Image Transformer, and Class-attention Image Transformer models