# APPLICATION FOR ETHICAL APPROVAL

**In order for research to result in benefit and minimise risk of harm, it must be conducted ethically.**

The University follows the OECD Frascati manual definition of **research activity**: "creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications". As such this covers activities undertaken by members of staff, postgraduate research students, and both taught postgraduate and undergraduate students working on dissertations/projects.

The individual undertaking the research activity is known as the "principal researcher".

**This form must be completed and approved prior to undertaking any research activity.**

## SECTION A: About You (Principal Researcher)

| | | |
|---|---|---|
| 1 | Full Name: | MD MAHADI HASAN |
| 2 | Student Number: | 2505591 |
| 3 | Email address: | 2505591@student.uwtsd.ac.uk |
| 4 | Programme of Study: | MSc in Software Engineering and AI |
| 5 | Supervisor: | |

## SECTION B: Internal and External Ethical Guidance Materials

| | | |
|---|---|---|
| | Please list the core ethical guidance documents that have been referred to during the completion of this form (including any discipline-specific codes of research ethics, location-specific dodes of research ethics, and also any specific ethical guidance relating to the proposed methodology). Please tick to confirm that your research proposal adheres to these codes and guidelines. You may add rows to this table if needed. | |
| 1 | **UWTSD Research Ethics & Integrity Code of Practice** | ☒ |
| 2 | **UWTSD Research Data Management Policy** | ☒ |
| 3 | | ☐ |
| 4 | | ☐ |

## SECTION C: Details of Research Activity

| | | | | |
|---|---|---|---|---|
| 1 | Indicative title: | Prediction of Early Gestational Diabetes Mellitus Using Machine Learning Model | | |
| 2 | Proposed start date: | 01-06-2026 | Proposed end date: | 25-09-2026 |
| | **Introduction to the Research (maximum 300 words in each section)**<br>**Ensure that you write for a <u>Non-Specialist Audience</u> when outlining your response to the three points below:**<br>• *Purpose of Research Activity*<br>• *Proposed Research Question*<br>• *Aims of Research Activity* | | | |

| | |
|---|---|
| | • *Objectives of Research Activity*<br>Demonstrate, briefly, how **Existing Research** has informed the proposed activity and explain<br>• *What the research activity will add to the body of knowledge*<br>• *How it addresses an area of importance.* |
| 3 | **Purpose of Research Activity**<br>Gestational diabetes mellitus (GDM) is a type of diabetes which develops during pregnancy, and which complicates the health of both mother and child. When not detected in the initial stages of pregnancy, it predisposes the mother to developing Type 2 diabetes in the long-term and is linked to complications in the baby. Nowadays, GDM is systematically diagnosed with the help of a glucose tolerance test performed between the 24th and 28th weeks of pregnancy a time when earlier intervention could have been made. This study applies machine learning (a form of computer analysis that identifies trends in data) to make earlier predictions of the risk of GDM during pregnancy based on routine clinical measurements of the predictive variables, including blood glucose levels, body mass index (BMI), age, and number of previous pregnancies. Three machine learning algorithms are compared to find which one is most accurate and understandable by healthcare professionals. The dataset used in the study is the publicly available and completely anonymised PIMA Indians Diabetes Dataset, sourced by the UCI Machine Learning Repository that contains 768 historical patient records. It does not collect any new patient data and does not recruit any human participants. The study also uses a method known as SHAP (SHapley Additive exPlanations) to explain which clinical factors have the largest impact on each prediction, thus the model outputs can be interpreted by the clinical staff. The result is a validated prediction instrument that may one day in the future be useful in supporting earlier and more focused antenatal screening.<br><br>(this box should expand as you type) |
| 4 | **Research Question**<br>Which machine learning classifier — Logistic Regression, Random Forest, or XGBoost — provides the most accurate and clinically interpretable prediction of gestational diabetes mellitus risk using routine early-pregnancy clinical data, and which individual clinical features most strongly influence prediction outcomes?<br><br>(this box should expand as you type) |
| 5 | **Aims of Research Activity**<br>Number your aims and link them to your objectives<br>**Aim:** To develop, train, and comparatively evaluate a machine learning-based prediction model for early gestational diabetes mellitus using clinical and demographic features, with the objective of identifying the most accurate and interpretable classifier suitable for early clinical risk stratification.<br>This aim is addressed through four linked objectives (see C6 below). Existing research has shown that Random Forest and XGBoost classifiers can achieve AUC values of 0.80–0.94 for GDM prediction, and that explainability frameworks such as SHAP are increasingly required for responsible clinical AI deployment. However, studies consistently identify class imbalance, over-reliance on the PIMA dataset without appropriate preprocessing, and the absence of integrated explainability as unresolved gaps. This study addresses those gaps by combining principled SMOTE-based class balancing, systematic multi-algorithm comparison, and SHAP feature analysis within a reproducible Design Science Research framework.<br><br>(this box should expand as you type) |
| 6 | **Objectives of Research Activity**<br><br>**O1:** To critically analyze existing machine learning approaches to GDM prediction through a systematic review of published literature (2022–2026), identifying methodological trends and research gaps.<br>**O2:** To design and develop an ML prediction artefact comprising Logistic Regression, Random Forest, and XGBoost classifiers, trained on preprocessed and SMOTE-balanced data from the PIMA Indians Diabetes Dataset.<br>**O3:** To evaluate the developed models using standard performance metrics (Accuracy, F1-score, AUC-ROC, Precision, Recall) and to apply SHAP analysis to interpret individual feature contributions to each model's predictions.<br>**O4:** To compare model results and identify the optimal GDM prediction classifier, explicitly assessing the trade-off between predictive accuracy and clinical interpretability. |

| | |
|---|---|
| | (this box should expand as you type) |
| | **Proposed data collection methods (maximum 600 words)**<br>Provide a brief summary of all the methods that **may** be used in the research activity to collect data, making it clear what specific techniques may be used. If methods other than those listed in this section are deemed appropriate later, additional ethical approval for those methods will be needed. You do not need to justify the methods here, but should instead describe how you intend to collect the data necessary for you to complete your project. |
| 7 | This should describe how you intend to collect data. It should not include a discussion of the theoretical basis for your data collection methods. Please note, that if you intend to collect any audio/video recordings of interviews with participants then these will be are classified as Personal Data under GDPR/DPA2018. If you intend to use these then note this in section H.<br>In this research, it is not required that the researcher should collect primary data by using human subjects. The only source of data is the PIMA Indians Diabetes Dataset, a publicly available, fully anonymised secondary dataset sourced in the UCI Machine Learning Repository. The dataset includes 768 historical patient records that include eight clinical features (number of pregnancies, plasma glucose, blood pressure, triceps skinfold thickness, serum insulin, BMI, diabetes pedigree function, and age) and a binary outcome label (diabetes positive/negative). The dataset does not include any personally identifiable data. The data is downloaded directly off the UCI repository and done entirely in a Google Collab Python notebook environment. To overcome class imbalance, preprocessing steps include median imputation of physiologically implausible zero values, Min-Max normalisation, and SMOTE oversampling. All stages do not use any audio, video, questionnaires, interviews, or observations.<br>(this box should expand as you type) |

**SECTION D: Scope of Research Activity**

| | **Will the research activity include:** | **YES** | **NO** |
|---|---|---|---|
| 1 | Use of a questionnaire or similar research instrument? | ☐ | ☒ |
| 2 | Use of interviews? | ☐ | ☒ |
| 3 | Use of focus groups? | ☐ | ☒ |
| 4 | Use of participant diaries? | ☐ | ☒ |
| 5 | Use of video or audio recording? | ☐ | ☒ |
| 6 | Use of computer-generated log files? | ☐ | ☒ |
| 7 | Participant observation with their knowledge? | ☐ | ☒ |
| 8 | Participant observation without their knowledge? | ☐ | ☒ |
| 9 | Access to personal or confidential information without the participants' specific consent? | ☐ | ☒ |
| 10 | Administration of any questions, test stimuli, presentation that may be experienced as physically, mentally or emotionally harmful / offensive? | ☐ | ☒ |
| 11 | Performance of any acts which may cause embarrassment or affect self-esteem? | ☐ | ☒ |
| 12 | Investigation of participants involved in illegal activities? | ☐ | ☒ |
| 13 | Use of procedures that involve deception? | ☐ | ☒ |
| 14 | Administration of any substance, agent or placebo? | ☐ | ☒ |
| 15 | Working with live vertebrate animals? | ☐ | ☒ |
| 16 | Procedures that may have a negative impact on the environment? | ☐ | ☒ |

| | | | |
|---|---|---|---|
| 17 | Other primary data collection methods. Please indicate the type of data collection method(s) below. | | |
| | Details of any other primary data collection method:<br><br>(this box should expand as you type) | ☐ | ☒ |

If you have ticked NO to every question then the research activity is (ethically) low risk and you may skip section E and continue to section F.

If YES to any question, then no research activity should be undertaken until full ethical approval has been obtained.

**SECTION E: Intended Participants**

| | **Who are the intended participants:** | **YES** | **NO** |
|---|---|---|---|
| 1 | Students or staff at the University? | ☐ | ☐ |
| 2 | Adults (over the age of 18 and competent to give consent)? | ☐ | ☐ |
| 3 | Vulnerable adults? | ☐ | ☐ |
| 4 | Children and Young People under the age of 18? (Consent from Parent, Carer or Guardian will be required) | ☐ | ☐ |
| 5 | Prisoners? | ☐ | ☐ |
| 6 | Young offenders? | ☐ | ☐ |
| 7 | Those who could be considered to have a particularly dependent relationship with the investigator or a gatekeeper? | ☐ | ☐ |
| 8 | People engaged in illegal activities? | ☐ | ☐ |
| 9 | Others. Please indicate the participants below, and specifically any group who may be unable to give consent. | | |
| | Details of any other participant groups:<br>Complete this only if your participants cannot give consent. This includes animals<br><br>(this box should expand as you type) | ☐ | ☐ |

| | **Participant numbers and source**<br>Provide an estimate of the expected number of participants. How will you identify participants and how will they be recruited? | |
|---|---|---|
| 10 | How many participants are expected? | Ballpark figures are fine, but make sure that you explain how you will identify and contact your participants.<br><br>*(this box should expand as you type)* |
| 11 | Who will the participants be? | *(this box should expand as you type)* |
| 12 | How will you identify the participants? | *(this box should expand as you type)* |

| | **Information for participants:** | **YES** | **NO** | **N/A** |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| 13 | Will you describe the main research procedures to participants in advance, so that they are informed about what to expect? | ☐ | ☐ | ☐ |
| 14 | Will you tell participants that their participation is voluntary? | ☐ | ☐ | ☐ |
| 15 | Will you obtain written consent for participation? | ☐ | ☐ | ☐ |
| 16 | Will you explain to participants that refusal to participate in the research will not affect their treatment or education (if relevant)? | ☐ | ☐ | ☐ |
| 17 | If the research is observational, will you ask participants for their consent to being observed? | ☐ | ☐ | ☐ |
| 18 | Will you tell participants that they may withdraw from the research at any time and for any reason? | ☐ | ☐ | ☐ |
| 19 | With questionnaires, will you give participants the option of omitting questions they do not want to answer? | ☐ | ☐ | ☐ |
| 20 | Will you tell participants that their data will be treated with full confidentiality and that, if published, it will not be identifiable as theirs? | ☐ | ☐ | ☐ |
| 21 | Will you debrief participants at the end of their participation, in a way appropriate to the type of research undertaken? | ☐ | ☐ | ☐ |
| 22 | If NO to any of above questions, please give an explanation | | | |
| | You should be able to tick YES for all of these questions. If not, then explain why not in this box.<br><br>*(this box should expand as you type)* | | | |

| | **Information for participants:** | **YES** | **NO** | **N/A** |
|---|---|---|---|---|
| 24 | Will participants be paid? | ☐ | ☐ | ☐ |
| 25 | Is specialist electrical or other equipment to be used with participants? | ☐ | ☐ | ☐ |
| 26 | Are there any financial or other interests to the investigator or University arising from this study? | ☐ | ☐ | ☐ |
| 27 | Will the research activity involve deliberately misleading participants in any way, or the partial or full concealment of the specific study aims? | ☐ | ☐ | ☐ |
| 28 | If YES to any question, please provide full details | | | |
| | You should be able to tick NO for most of these questions. For any cases that you have ticked YES then provide details in this box. If you are using cameras/voice recorders to record interviews then please state that in this box.<br><br>*(this box should expand as you type)* | | | |

## SECTION F: Anticipated Risks

| | **Outline any anticipated risks that may adversely affect any of the participants, the researchers and/or the University, and the steps that will be taken to address them.** | |
|---|---|---|
| 1 | **Risks to participants**<br>For example: sector-specific health & safety, emotional distress, financial disclosure, physical harm, transfer of personal data, sensitive organisational information. If you have identified in section D that there are no participants then enter N/A and go skip to question 3. | |
| | Risk to participants:<br>There are always risks. Do not write N/A unless you have no participants.<br>This study uses only a publicly available, fully anonymised secondary dataset. No human participants are recruited or involved in any stage of the research.<br>*(this box should expand as you type)* | *How you will mitigate the risk to participants:*<br><br>*(this box should expand as you type)* |
| 2 | If research activity may include sensitive, embarrassing or upsetting topics (e.g. sexual activity, drug use) or issues likely to disclose information requiring further action (e.g. criminal activity), give | |

| | | |
|---|---|---|
| | details of the procedures to deal with these issues, including any support/advice (e.g. helpline numbers) to be offered to participants. Note that where applicable, consent procedures should make it clear that if something potentially or actually illegal is discovered in the course of a project, it may need to be disclosed to the proper authorities | |
| | Not applicable. No participants are involved and no primary data collection takes place<br><br>*(this box should expand as you type)* | |
| 3 | **Risks to the investigator**<br>For example: personal health & safety, physical harm, emotional distress, risk of accusation of harm/impropriety, conflict of interest | |
| | Risk to the investigator:<br>There are always risks. Do not write NA.<br>Academic integrity risk — potential for the ML model to produce misleading performance results if preprocessing errors occur<br>Data handling risk — inadvertent use of a non-anonymised dataset version<br><br>*(this box should expand as you type)* | *How you will mitigate the risk to the investigator:*<br>All preprocessing steps (imputation, normalisation, SMOTE) are documented and reproducible within a version-controlled Google Collab notebook. Results are reported with appropriate limitations.<br><br>The dataset is sourced only from the official UCI ML Repository and verified as containing no personal identifiers before use<br>*(this box should expand as you type)* |
| 4 | **University/institutional risks**<br>For example: adverse publicity, financial loss, data protection | |
| | Risk to the University:<br>There are always risks. Do not write NA.<br>Reputational risk if ML outputs are presented without appropriate caveats and interpreted as clinical recommendation<br><br>*(this box should expand as you type)* | *How you will mitigate the risk to the University:*<br>The proposal and report clearly state that the artefact is a research prototype requiring external clinical validation before any deployment consideration.<br><br>*(this box should expand as you type)* |
| 5 | **Environmental risks**<br>For example: accidental spillage of pollutants, damage to local ecosystems | |
| | Risk to the environment:<br>Minor energy consumption from cloud-based computation (Google Collab servers)<br><br>*(this box should expand as you type)* | *How you will mitigate the risk to environment:*<br>The study uses classical ML classifiers on a small dataset (n=768), requiring minimal computational resource. No GPU-intensive deep learning is used. Environmental impact is negligible.<br><br>*(this box should expand as you type)* |

**SECTION G: Feedback, Consent and Confidentiality**

If you have identified in section D that there are no participants then enter skip this section and continue to section H.

| | |
|---|---|
| 1 | **Feedback**<br>What de-briefing and feedback will be provided to participants, how will this be done and when? |
| | You don't need to email your participants with your final report. A good alternative is to set up an email address that they will be able to contact for further details or results.<br><br>*(this box should expand as you type)* |
| 2 | **Informed consent**<br>Describe the arrangements to inform potential participants, before providing consent, of what is involved in participating. Describe the arrangements for participants to provide full consent before data collection begins. If gaining consent in this way is inappropriate, explain how consent will be obtained and recorded in accordance with prevailing data protection legislation. |

| | |
|---|---|
| | If you are using a paper questionnaire then you should have the participants sign an appropriate consent form. These forms will count as personal data and should be noted as such in section J.<br>If you are using an online questionnaire, then you should have a screen before the questions start that acts as a consent form, informing participants that by clicking on the NEXT button they are providing consent.<br><br>*(this box should expand as you type)* |
| 3 | **Confidentiality / Anonymity**<br>Set out how anonymity of participants and confidentiality will be ensured in any outputs. If anonymity is not being offered, explain why this is the case. |
| | Do not collect names unless you really need them. Do not name participants or organisations in any research publications (including the thesis) without their explicit permission.<br><br>*(this box should expand as you type)* |

**SECTION H: Data Protection and Storage**

| | Does the research activity involve personal data (as defined by the General Data Protection Regulation 2016 "GDPR" and the Data Protection Act 2018 "DPA")? | **YES** | **NO** |
|---|---|---|---|
| 1 | ***"Personal data"*** *means any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Any video or audio recordings of participants is considered to be personal data.* | ☐ | ☐ |
| | If YES, provide a description of the data and explain why this data needs to be collected: | | |
| 2 | This includes audio/video data of participants, but can also include IP addresses and usernames. Names, addresses and emails also count, as do consent forms.<br><br>*(this box should expand as you type)* | | |
| | Does it involve special category data (as defined by the GDPR)? | **YES** | **NO** |
| 3 | ***"Special category data"*** *means sensitive personal data consisting of information as to the data subjects' –*<br>*(a) racial or ethnic origin,*<br>*(b) political opinions,*<br>*(c ) religious beliefs or other beliefs of a similar nature,*<br>*(d) membership of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992),*<br>*(e) physical or mental health or condition,*<br>*(f) sexual life,*<br>*(g) genetics,*<br>*(h) biometric data (as used for ID purposes),* | ☐ | ☐ |
| | If YES, provide a description of the special category data and explain why this data needs to be collected: | | |
| 4 | What counts as 'sensitive' will differ between cultures. Any information on behaviour that is not in accordance with cultural norms would count as sensitive personal data.<br><br>*(this box should expand as you type)* | | |

| | **Will data from the research activity (collected data, drafts of the thesis, or materials for publication) be stored in any of the following ways?** | **YES** | **NO** |
|---|---|---|---|
| 5 | Manual files (i.e. in paper form)? | ☐ | ☒ |

| | | | |
|---|---|---|---|
| 6 | University computers? | ☐ | ☒ |
| 7 | Private company computers? | ☒ | ☐ |
| 8 | Home or other personal computers? | ☒ | ☐ |
| 9 | Laptop computers/ CDs/ Portable disk-drives/ memory sticks? | ☐ | ☒ |
| 10 | "Cloud" storage or websites? | ☒ | ☐ |
| 11 | Other – specify: | ☐ | ☒ |
| 12 | For all stored data, explain the measures in place to ensure the security of the data collected, data confidentiality, including details of backup procedures, password protection, encryption, anonymisation and pseudonymisation: | | |
| | If possible, save your data on computers that are secure and regularly backed up. Many cloud services only provide GDPR-compliant storage for business customers. An example of suitable text is given below.<br><br>*All data will be kept in password protected cloud storage on the University Office 365 system which will not be shared. Audio/visual data will be transcribed and would be shown to participants to check accuracy of reporting. Any USB sticks used to store or transfer data will be password protected. All participants will be given a unique identifier to ensure confidentiality and this list will be kept securely in the password protected folder.*<br>The publicly accessible, anonymised PIMA data and Python notebook files with no personal data comprise all the research data. All thesis drafts and notebooks are stored in password-protected University OneDrive (Microsoft Office 365) cloud storage and are not shared with third parties. The use of Google Collab is only applied to calculate; no personal information is uploaded to it. Personal devices have local copies that are stored in a password-protected folder. All information will be saved until the submission of the project and be removed in compliance with the university policy.<br><br>*(this box should expand as you type)* | | |

| | | | |
|---|---|---|---|
| | **Data Protection** | | |
| | Will the research activity involve any of the following activities: | **YES** | **NO** |
| 13 | Electronic transfer of data in any form? | ☒ | ☐ |
| 14 | Sharing of data with others at the University outside of the immediate research team? | ☒ | ☐ |
| 15 | Sharing of data with other organisations? | ☐ | ☒ |
| 16 | Export of data outside the UK or importing of data from outside the UK? | ☒ | ☐ |
| 17 | Use of personal addresses, postcodes, faxes, emails or telephone numbers? | ☐ | ☒ |
| 18 | Publication of data that might allow identification of individuals? | ☐ | ☒ |
| 19 | If YES to any question, please provide full details, explaining how this will be conducted in accordance with the GDPR and Data Protection Act (2018) (and/or any international equivalent): | | |
| | This includes data such as drafts of your thesis as well as experimental or survey data. An example of suitable text is given below.<br><br>*All data will be encrypted and kept in password protected cloud storage on the University Office 365 system which will not be shared. Any USB sticks used to store or transfer data will be password protected. All data transfers will be encrypted and password protected. All participants will be given a unique identifier to ensure confidentiality and this list will be kept securely in the password protected folder. The data will be stored until the completion of the project and then deleted. In accordance with the DPA2018, participants will have the right to ask to see what data is held relating to them, and this data will be deleted immediately if the participant requests this, in which case the data will not be used in the project.* | | |

| | |
|---|---|
| | The PIMA data is obtained via the UCI Machine Learning Repository as a download using HTTPS. It does not entail any personal information. Thesis drafts submitted to the supervisor are saved in the password-protected University OneDrive system. Although US-based, Google Collab only processes anonymised public data and no personal or special category data is uploaded anywhere. Any transfer will be done with anonymised research data or academic draft documents only.<br>*(this box should expand as you type)* |
| 20 | List all who will have access to the data generated by the research activity: |
| | Normally the principal researcher, possibly also the supervisor and, if the project has an industrial partner, a representative of that partner. Possibly also external examiner or second marker?<br>Principal researcher (Md Mahadi Hasan) and module supervisor<br>*(this box should expand as you type)* |
| 21 | List who will have control of, and act as custodian(s) for, data generated by the research activity: |
| | Usually the principal researcher.<br>Principal researcher (Md Mahadi Hasan).<br>*(this box should expand as you type)* |
| 22 | Give details of data storage arrangements, including security measures in place to protect the data, where data will be stored, how long for, and in what form. |
| | *All data will be encrypted and kept in password protected cloud storage on the University Office 365 system which will not be shared. Any USB sticks used to store or transfer data will be password protected, and will be reformatted at the end of the project in order to destroy the data. The data will be stored until the completion of the project and then deleted.*<br>All information is saved in a secure University OneDrive cloud storage with a password. Publicly available PIMA dataset does not include any personal information and does not necessitate any special protective measures beyond the use of passwords. Thesis drafts are kept in the same secure facility. Any files will be destroyed after the project is completed and submitted. justification and tool description:<br>*(this box should expand as you type)* |
| 22 | Confirm that you have read the UWTSD guidance on data management (see https://www.uwtsd.ac.uk/library/research-data-management/) ☐ |
| 23 | Confirm that you are aware that you need to keep all data until after your research has completed or the end of your funding ☐ |

**SECTION I: Use of AI tools**

| **Use of generative AI and other AI-based tools**<br>There are a number of ethical issues relating to the use of generative AI and other AI-based tools, including data security, output bias, and output errors. In general you should assume that any generative AI or other AI-based tool that you pay to use is more secure and accurate than one which is free to use.<br><br>Will the research use AI tools to do any of the following: | | **YES** | **NO** | **N/A** |
|---|---|---|---|---|
| 1 | Capture data? | ☐ | ☒ | ☐ |
| 2 | Generate and/or edit text? | ☐ | ☒ | ☐ |
| 3 | Generate graphics? | ☐ | ☒ | ☐ |
| 4 | Generate video? | ☐ | ☒ | ☐ |
| 5 | Generate sound/music? | ☐ | ☒ | ☐ |
| 6 | Generate code? | ☐ | ☒ | ☐ |
| 7 | Generate synthetic data? | ☒ | ☐ | ☐ |
| 8 | Analyse data? | ☒ | ☐ | ☐ |

| | | | | |
|---|---|---|---|---|
| 9 | Analyse personal or sensitive data? | ☐ | ☒ | ☐ |
| 10 | Classify data? | ☒ | ☐ | ☐ |
| 11 | Do you intend to use the free version of any AI tools? | ☒ | ☐ | ☐ |
| If you have answered **YES** to any of questions O1-O11, please provide justification detailing why this is necessary for the research. | | | | |
| 12 | State which AI tool or tools you will be using, and where they are based (or store their data – you may need to check their webpages or user agreement for this). If you are planning to use Copilot then say if you're using the UWTSD version of it or the public version. You should also show how you are complying with current UWTSD guidance/policy of the use of AI.<br>The machine learning models employed in this paper (Logistic Regression, Random Forest, and XGBoost, implemented using Scikit-learn and the XGBoost Python library) are AI tools, which analyse and categorize data. Synthetic training samples are produced using SMOTE (part of the Imbalanced-learn library) to generate the synthetic samples. SHAP is used to analyse feature contributions to model predictions. Each of the tools is an open-source Python library that does not store any data proprietarily. It is computed in Google Collab (Google LLC, USA - data computed on the cloud servers of Google). The publicly available and anonymised PIMA Indians Diabetes Dataset are the only data uploaded to Collab; no personal or sensitive information is processed by any AI tool. The requirement of these tools is based on the research objective, which is to make a comparative evaluation of the ML classifiers in GDM prediction. The dataset is documented to have an imbalance in the number of classes. SMOTE is mandated to overcome the number of classes imbalance that is documented in the dataset. SHAP must produce explanations of features that are clinically interpretable. The use of these tools follows UWTSD AI guidance since no personal data is processed, outputs are clearly attributed to algorithmic analysis, and all the use of these tools is transparently reported in the research methods.<br><br>*(this box should expand as you type)* | | | |

**SECTION J: Declaration**

| | | | |
|---|---|---|---|
| | The information which I have provided is correct and complete to the best of my knowledge. I have attempted to identify any risks and issues related to the research activity and acknowledge my obligations and the rights of the participants.<br><br>In submitting this application I hereby confirm that I undertake to ensure that the above named research activity will meet the University's Research Ethics and Integrity Code of Practice which is published on the website: https://www.uwtsd.ac.uk/research/research-ethics/ | | |
| 1 | **Signature of applicant:** | MD MAHADI HASAN | **Date:08-05-2026** |
| 2 | Supervisor: | | **Date:** |
| 3 | Signature: | Mahadi | |