Problem Definition and Design Thinking (AI-Based Diabetes Prediction System)

**Introduction:**

Diabetes is a significant global health challenge, and early prediction is essential for effective management and prevention. In this project, we aim to develop an AI-Based Diabetes Prediction System utilising advanced techniques, including Convolutional Neural Networks (CNN2D) with Particle Swarm Optimization (PSO) for feature selection. Additionally, we will employ ensemble learning, specifically the tabPFN model, to further enhance prediction accuracy. This system will empower individuals and healthcare professionals with accurate diabetes risk assessments.

**Problem Definition:**

The primary goal of this project is to create an AI-Based Diabetes Prediction System with the following key components:

**Data Collection:**

* Gather a comprehensive dataset comprising patient demographics, medical history, lifestyle factors, and biomarker measurements related to diabetes.
* Explore data sources such as electronic health records, public health datasets, and research studies.

**Data Preprocessing:**

* Prepare and clean the acquired data, addressing missing values, outliers, and inconsistencies.
* Normalise or scale the data to ensure uniformity for modelling.

**Feature Selection with PSO:**

* Utilize Particle Swarm Optimization (PSO) to optimize feature selection, identifying the most relevant features for diabetes prediction.
* Incorporate domain knowledge to guide the feature selection process.

**CNN2D Model:**

* Implement a Convolutional Neural Network (CNN2D) architecture tailored for processing multi-dimensional data.
* Train the CNN2D model on the selected features for diabetes risk prediction.

**Ensemble Learning with tabPFN:**

* Develop an ensemble model using the tabPFN algorithm to combine the predictions from the CNN2D model and possibly other complementary models.
* Evaluate the ensemble model's ability to improve prediction accuracy.

**Model Training and Evaluation:**

* Split the dataset into training and testing sets for model training and validation.
* Fine-tune hyperparameters for both the CNN2D model and the ensemble model.
* Evaluate the models using appropriate metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

**Significance and Impact:**

The AI-Based Diabetes Prediction System developed in this project holds immense potential to revolutionize diabetes risk assessment and prevention. Its impact includes:

Early Identification: Enabling early identification of individuals at risk of diabetes, facilitating timely intervention and lifestyle modifications.

Clinical Decision Support: Assisting healthcare providers in identifying high-risk patients who may require closer monitoring and intervention.

Research Advancement: Contributing to the advancement of diabetes research by identifying key risk factors and predictive biomarkers.

The system's beneficiaries encompass individuals concerned about their diabetes risk, healthcare professionals, researchers in diabetes and preventive medicine, and public health organizations.

**Conclusion:**

In summary, our project aims to develop an AI-Based Diabetes Prediction System utilizing cutting-edge techniques, including CNN2D with PSO feature selection and ensemble learning with tabPFN. This comprehensive approach will ensure the creation of a powerful tool for accurate diabetes risk assessment, with the potential to significantly impact public health and individual well-being.