Principal Investigator	Dr. Marcos Lacasa-Cazcarra
Institution	Universidad Internacional de La Rioja (UNIR)
Project Summary	This submission requests data-only access to two mapMECFS studies: CFI, to perform external validation and cross-cohort harmonization of patient-reported outcome measures (PROMs) in ME/CFS. The project focuses on item-level ("raw") questionnaire responses and their official derived scores, with SF-36 as the primary anchor instrument and PSQI as a key target instrument. The work consists exclusively of secondary analysis of de-identified data (no participant contact, no interventions). Biospecimens are not requested in this phase. Primary objectives (CFI data-only) 1. External validation across independent cohorts: Validate whether PROM patterns and relationships observed in our local cohort (UNIR/VHIR) generalize to larger independent cohorts (CFI). 2. SF-36 anchor validation: Assess cross-cohort consistency of SF-36 item-level distributions and derived scores, and quantify comparability between cohorts. 3. PSQI replication and linkage: Using PSQI item-level responses and official scores, evaluate whether sleep-related outcomes replicate and how they relate to SF-36 across cohorts. 4. Scoring reproducibility: Confirm that analyses based on raw item responses reproduce the official derived scores (SF-36, PSQI and other instruments where available), supporting harmonized cross-study analysis. Secondary objectives 5. Secondary interpretation using additional PROM scores: Use available derived scores for additional fatigue, pain, psychological, and symptom instruments to support secondary interpretation and mapping to our local measures (HADS/HAD, FIS-40, COMPASS-31). 6. Exploratory PROM-based subgrouping: Explore whether stable PROM-derived subgroups can be detected consistently across CFI, informing later phases integrating objective measures (e.g., CPET) when available.
Project Non-Technical Summary	We are requesting access to de-identified questionnaire data to confirm that findings from our ME/CFS research replicate in larger independent cohorts. We will focus on two widely used questionnaires—SF-36 (quality of life) and PSQI (sleep)—using both the original responses and the official scores to ensure results are reliable and comparable across studies. This work will help identify consistent patterns in patient-reported outcomes across cohorts and improve reproducibility, without contacting participants and without requesting biospecimens in this phase.
Data Access Granted	CFI and MCAM
Project Timeline	Approval Date: 3/31/2026 Expected Close-out Date: 3/31/2027

Principal Investigator	Dr. Ashis Kumer Biswas
Institution	University of Colorado Denver
Project Summary	The project will combine datasets from multiple studies that have tested different potential biological indicators (protein levels, RNA, Genes, lipid levels, vitamin levels, enzyme levels) of ME/CFS as well as datasets from studies that have analyzed different presentations of ME/CFS with a focus on symptoms and/or comorbidities. The data will be pre-processed by applying a statistical method known as Principal Component Analysis to determine which potential indicators (features) have the most predictive potential of ME/CFS, selecting enough features to account for 95% of the data's variability. Then, the pre-processed data set will be split, 80% being used to train a simple neural network (1-3 hidden layers, n =< 5 neurons per layer) and 20% 8 being used to test the model to assess its performance. The model will use a sigmoid activation function at the hidden and output layers, conducting forward and backward passes for 5 epochs, or iterations. This process will be repeated with 10, 15, & 20 epochs to determine the optimal number of epochs. This process will then be repeated as needed to fine-tune the model's hyperparameters (activation function, number of layers, number of neurons per layer). A second convolutional neural network (CNN) will be trained with a larger dataset constructed from the same pre-processed dataset described previously. Said pre-processed dataset will be fed to a Conditional Generative Adversarial Network (CGAN) to generate artificial data in order to increase the sample size of the pre-processed dataset. This second, augmented dataset will be split with 80% used to train a CNN with 64 neurons per layer, 1-5 hidden layers and 20% used to test the CNN and evaluate its performance. Then the performance of both models' will be compared to determine the optimal approach to training a neural network classifier to attempt to detect the probability of ME/CFS in a patient. As previously mentioned, our CNN will require peer-review from medical researchers before a functional diagnostic tool can be developed.
Project Non-Technical Summary	The project will consist of compiling datasets from multiple studies investigating potential biological indicators of ME/CFS as well as studies evaluating the comorbidities and symptomatic presentations of ME/CFS. Due to the rarity of ME/CFS, sample sizes are expected to be small (ranging from 20-200 patients per study). Statistical methods will be used to determine which of the potential indicators, comorbidities, or symptomatic presentations of ME/CFS are more indicative of the presence of ME/CFS. Following the statistical analysis to reduce the number of indicators on the original datasets, two neural networks (computer models relying on deep learning techniques) will be trained. The first one will use the reduced dataset while the second one will use an "augmented" dataset that will be created by using generative models to analyze the data and generate

	artificial samples to increase its size. The performance of both models will be then compared to determine which approach was better at producing a computer model that can predict the probability of a patient having ME/CFS. This project aims to produce a computer model for educational purposes only. Further study involving medical doctors will be required before a real diagnostic tool can be developed.
Data Access Granted	MCAM Adult Longitudinal Study
Project Timeline	Approval Date: 5/1/2026 Expected Close-out Date: 5/1/2027