Performance of Machine Learning Models for Predicting High-Severity Symptoms in Multiple Sclerosis

Supplementary Materials

Contents

1 Data domains and modalities

The study was designed to collect data from the subjects via the MS Mosaic App, mobile phone, and wearable devices (if paired). The study was set up to collect data about various functional systems: gross motor function, autonomic, circadian rhythm, behavioural, mental, social, and cognitive changes, fine motor control, tacticle sensation, medication usage, and co-morbidity controls. Table 1 reports the functional tests, surveys, and physiological signals that the study had provision to record and categorizes them across functional domains and device. Table 1 also shows the sampling frequency of each datastream. However, not all modalities were filled by the participants and the ones recorded are marked by an asterisk at the end.

Table 1: Modalities of data recorded by the MS Mosaic App.

Domain	Device	Datastream	Sampling frequency
Gross Motor Function	MS Mosaic App	25-Foot walk task*	Bi-weekly
		Daily symptom survey*	Daily
	Phone	Accelerometer	Continuous
		Pedometer*	2-5 seconds
	Paired Devices	Accelerometer	Continuous while worn
		Pedometer*	2-5 seconds
		Stairs climbed	Event-triggered
		Stand hours	Hourly
		Workout sessions	Event-triggered
Autonomic	MS Mosaic App	Daily symptom survey*	Daily
	Paired Devices	Heart rate*	Seconds, dynamic
Circadian Rhythm	MS Mosaic App	Daily symptom survey	Daily
	Paired Devices	Sleep sensors*	Multiple
		Sleep summaries	Daily
Behavioral, Mental, Social, and Cognitive	MS Mosaic App	Spatial memory task*	Bi-weekly
		PASAT task*	Bi-weekly
		Self-efficacy survey*	Every three months, with first completed after enrollment
		Mindful minutes	Event-triggered
		Relapse survey*	Bi-weekly
		Daily symptom survey*	Daily
	Paired Devices	Breathe sessions	Event-triggered
Fine Motor Control	MS Mosaic App	Tapping Task	Bi-weekly
		9-Hole Peg Test*	Bi-weekly
Tactile Sensation	MS Mosaic App	Daily symptom survey*	Daily
Medication Usage	MS Mosaic App	Daily medication survey*	Daily
Comorbidity controls	MS Mosaic App	Height	Event-triggered
		Weight	Event-triggered
		BMI	Event-triggered
		VO2 max	Event-triggered
		Body temperature	Event-triggered
		Electrodermal activity	Event-triggered
		Menstrual cycle status	Event-triggered
		Hearing health	Event-triggered
		Nutrition	Event-triggered
		Number of times fallen	Event-triggered

2 Co-morbidities

Table 2 reports the co-morbidities reported by the participants in the development and blind test cohort.

Table 2: Co-morbidities in the development and blind test cohorts.

Characteristic	Development cohort	Blind test cohort
Medical history, No. (%)		
MS in family	109 (19.22)	37 (25.34)
Vitamin D deficiency	253 (44.62)	65 (44.52)
Hypertension	94 (16.57)	23 (15.75)
Thyroid disease	71 (12.52)	11 (7.53)
Cancer	29 (5.11)	5 (3.42)
Diabetes	27 (4.76)	7 (4.79)
Seizure or epilepsy	26 (4.58)	6 (4.10)
Other heart condition	32 (5.64)	7 (4.79)
Lyme disease	12 (2.12)	4 (2.74)
Neuromyelitis optica	10 (1.76)	1 (0.68)
Rheumatoid arthritis	10 (1.76)	3 (2.05)
Hepatitis	10 (1.76)	1 (0.68)
Meningitis	9 (1.59)	5 (3.42)
Lupus	7 (1.23)	0 (0.0)
Stroke	4 (0.71)	2 (1.37)
Sjogren's syndrome	4 (0.71)	2 (1.37)
Sarcoidosis	3 (0.53)	0 (0.0)
Tuberculosis	2 (0.35)	0 (0.0)
Heart attack	2 (0.35)	0 (0.0)
Muscle disease	2 (0.35)	0 (0.0)

3 Receiver operating characteristic, precision-recall, and calibration curves

Figure 1 shows the receiver operating characteristic, precision recall, and calibratuin curves obtained by the machine learning models for the symptom prediction tasks. The curves further demonstrate the superiority of GBC compared to other models.

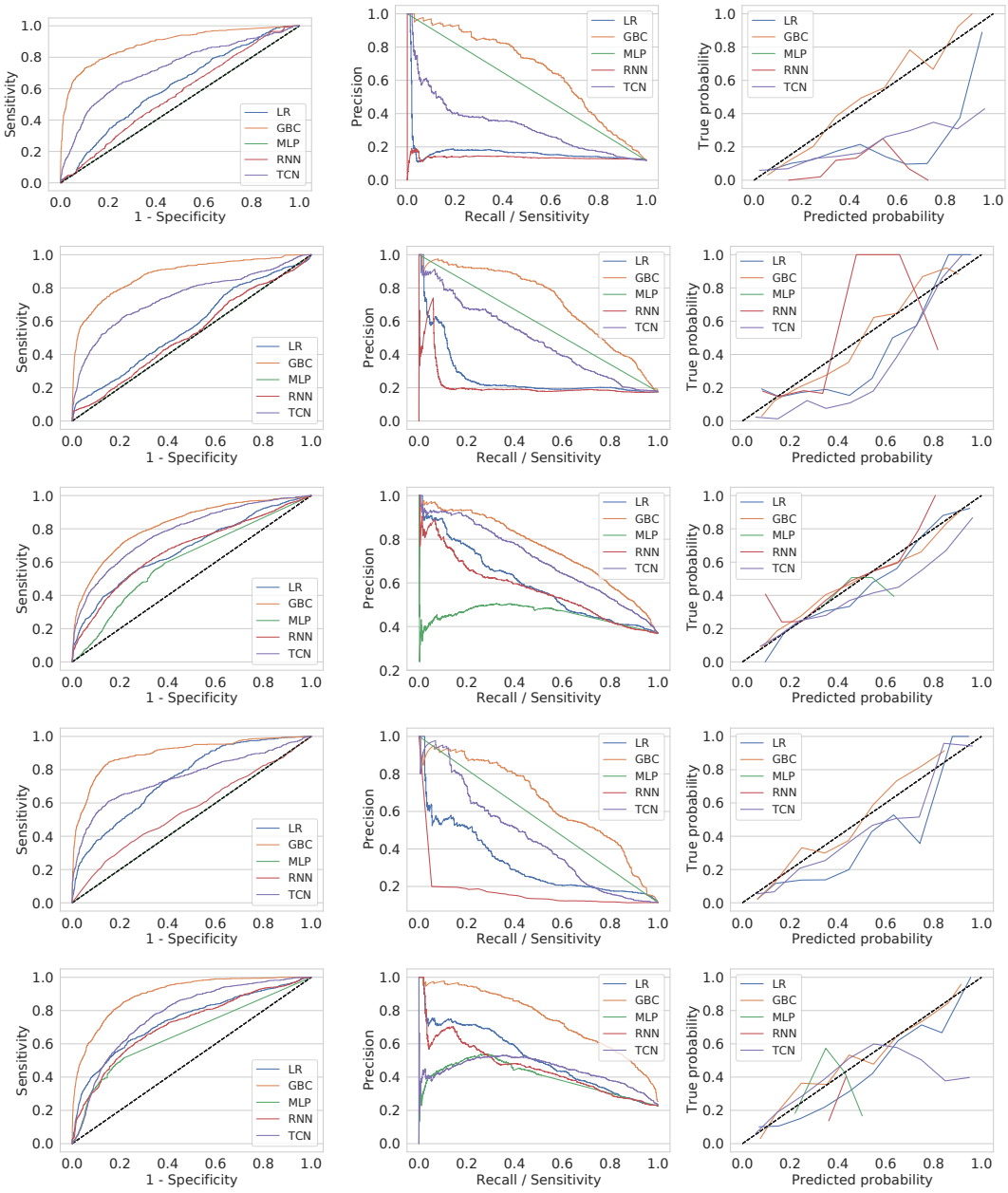

Figure 1: Receiver operating characteristic curves (left column), precision recall curves (middle column), and calibration curves (right column) for machine learning models (gradient boosting classifier, logistic regression, multi-layer perceptron, recurrent neural network, and temporal convolutional network) for predicting whether the median value of a user-reported symptom will be of high-severity (>2) in the next three months. The symptoms considered are (A) cramps, (B) depression or anxiety, (C) fatigue, (D) sensory disturbance, and (E) walking instability or coordination problems respectively.

4 Data processing and modelling pipeline

We developed a general data processing, modelling and evaluation pipeline which allows us to reuse the pipeline for a number of downstream prediction tasks and enables generalisation across datasets we may encounter in future [1]. A high-level overview of the pipeline is shown in Figure 2 and described below.

The raw data is taken through a series of steps to create the Subject representation shown in Table 3. The Subject representation consists of a predefined yet expandable set of fields which can used to map a wide range of datasets. The fields in Subject are selected such that they may be applied generally to various healthcare datasets as well as MS datasets (clinical and at-home). Additionally, the representation can be enlarged to incorporate particular dataset-specific nuances.

Following the transformation of raw MS Mosaic dataset into the Subject representation, a Label Creator is run on all processed data and add labels to it before the data is translated into model input. For this study, we only added symptom-derived labels. The suggested pipeline is feature agnostic, meaning that it does not require any special pre-processing for any particular dataset. All model output is saved using a prediction format after training. The metrics pipeline then receives this, which computes performance metrics at both the population-level and for multiple clinically relevant subgroups.

The Prediction format, which is defined in Table 4, is straightforward in setup yet allows for the building of downstream metrics pipelines and cross-model comparisons. While it primarily focuses on time series in its current form (by using the timestamp column), we anticipate that it may also be applied to other types by only ignoring the time value. To make at-scale metrics computations easier, we decided to store a number of label targets and predictions at once. As a result, our pipeline computes a number of metrics simultaneously for each operation.

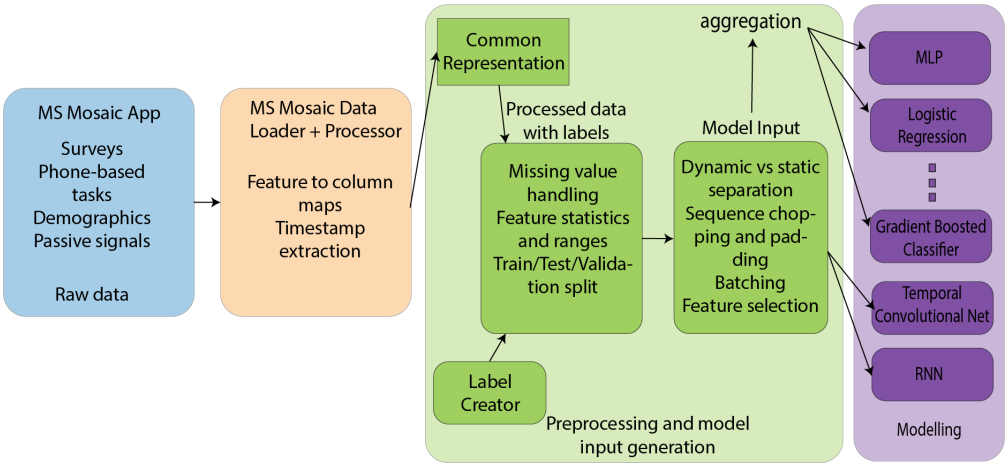

Figure 2: Data processing and modelling pipeline.

Table 3: Description of the common Subject representation.

Field	Type	Format	Description
Subject: Defines all the data provided for a single subject.			
subject_id	string	Optional	Unique ID for each subject.
subject_characteristics	SubjectCharacteristics	Optional	Defines the subject's characteristics.
medical_history	MedicalHistory	Optional	Defines the subject's medical history.
episodes	Episode	Repeated	A sequence of encounters corresponding to a single subject.
SubjectCharacteristics: Defines all the data provided for a single subject.			
sex	Sex.Enum	Optional	The subject's sex.
MedicalHistory: Defines a sequence of outpatient events recorded before the current system started recording events. These typically are recorded with the same timestamp even though they took place over longer periods of time.			
clinical_events	ClinicalEvent	Repeated	A clinical event.
Episode: An abstraction of an event sequence.			
clinical_event	ClinicalEvent	Repeated	A sequence of clinical events.
changing_characteristics	ChangingSubjectCharacteristics	optional	Defines the subject's changing characteristics.
clinical_trial	ClinicalTrial	optional	Details about the clinical trial the subject was part of (if applicable).
ChangingSubjectCharacteristics: Defines the subject's changing characteristics.			
age	float	optional	The subject's age.
gender	Gender.Enum	optional	The subject's gender
race	string	optional	The subject's race.
weight	float	optional	The subject's weight.
height	float	optional	The subject's height.
country	string	optional	The subject's country.
Sex.Enum: Defines the subject's sex information.			
FEMALE	0	-	Female sex.
MALE	1	-	Male sex.
Gender.Enum: Defines the subject's gender information.			
UNKNOWN	0	-	Unknown gender.
FEMALE	1	-	Female gender.
MALE	2	-	Male gender.
OTHER	3	-	Other gender.
ClinicalEvent: Defines a single event or a set of coinciding individual events that happen at the same time.			
timestamp	int64	optional	Timestamp of event. While for MSOAC, this corresponds to the day of the clinic visit, for Floodlight, this is the timestamp recorded by the smartphone.
resources	Resource	repeated	A list of specific clinical entries recorded at this timestamp.
classification_labels	map<string, int64>	required	Classification labels for prediction.
regression_labels	map<string, float>	required	Regression labels for prediction.
Resource: Describes the various types of resources that can be contained within ClinicalEvent.			
functional_test	FunctionalTest	optional	Functional assessment test data.
questionnaire	Questionnaire	optional	Questionnaires filled by the subjects.
generic_resource	GenericResource	optional	Generic resource to store dataset-specific intricacies.
FunctionalTest: Functional assessment test data recorded from the subject.			
name	string	optional	Name of performance outcome measure.
category	string	optional	Category of performance outcome measure.
response	NumericalResponse	optional	Numerical response recorded from subject.
Questionnaire: Questionnaire data recorded from the subject.			
name	string	optional	Name of performance outcome measure.
category	string	optional	Category of performance outcome measure.
response	QuestionnaireResponse	optional	Questionnaire response recorded from subject.
NumericalResponse: Numeric response converted to standardized unit.			
numerical_response_std_unit	float	optional	Numeric response converted to standardized unit.
std_unit	string	optional	Standardized unit. This unit was used for homogenizing the data.
QuestionnaireResponse: Responses recorded from questionnaires.			
text_response_orig	string	optional	Response in original text.
text_response_std	string	optional	Response in standardized text.
numeric_response_std	float	optional	Standardized numeric response.
categorical_response_std	string	optional	Standardized categorical response.An example entry is EDSS.

Table 4: Description of the common Prediction representation.

Field	Type	Description
Prediction: Defines the model output information.		
subject_id	string	Unique ID for each subject.
timestamp	int64	Timestamp of event. While for MSOAC, this corresponds to the day of the clinic visit, for Floodlight, this is the timestamp recorded by the smartphone.
label_targets	map<string, float>	A mapping from label name to the target value for the particular timestamp this is recorded for.
label_predictions	map<string, list<float>>	A mapping from label name to the predicted values for the particular timestamp this is recorded for. Multiple values are recorded to account for multi-class predictions.
subgroup_attributes	map<string, oneof<string, int, float>>	A mapping from subgroup name (such as Sex, or Age) to the exact value (e.g. Female, or 56).

5 List of features used by the machine learning models

Below we list the features used by the ML models categorized according to type. Among the modalities reported in Table 1, we chose only the ones which had more than a single data-point for the cohort selected for this study. Note that out of the surveys we only used recorded symptoms since our prediction labels were derived from symptom data. We did not use medication surveys there were around 700+ medications which would have led to an explosion of features. Furthermore, harmonizing the medications to create a smaller subset was beyond the scope of this work. We intend to use the medication data as features in future work.

- Demographics
  - Age
- Functional tests
  - Tapping Task: Left finger speed
  - Tapping Task: Number of left finger taps
  - Tapping Task: Right finger speed
  - Tapping Task: Number of right finger taps
  - 9-Hole Peg Test: Left hand remove time
  - 9-Hole Peg Test: Right hand remove time
  - 25-Foot walk task: Walking outbound time
  - 25-Foot walk task: Walking inbound time
  - PASAT task: Correct count
  - PASAT task: Series length
  - Spatial memory task: Score
- Symptom severity
  - Double vision
  - Cognitive changes or brain fog
  - Headache
  - Fatigue
  - Walking instability or coordination problems
  - Vertigo or dizziness
  - Weakness
  - Sensory disturbance
  - Bowel problems
  - Sexual problems

  - Bladder problems
  - Depression or anxiety
  - Cramps
  - Vision loss/blurred vision
  - Hearing changes
  - Slurred speech
  - Loss of taste
  - Difficulty swallowing
  - Facial weakness
- Passive signals
  - Daily steps
  - Daily hours of sleep
  - Daily distance walked
  - Heart rate

6 Suggestions for app-based study development

Based on the findings while running this study, we would like to make a set of recommendations for future studies:

1. Communication between all parties involved is key. As each field has its own specific language, making sure there is someone interfacing between the stakeholders is key. For example, app developers might not understand why a given user interface would not work for an MS patient, as they might not be aware of the symptom range one experiences.

2. If possible, use existing platforms for data collection rather than building everything from the ground up. It will reduce a lot of overhead associated with the set-up.

3. Be aware of hidden costs. There is a constant need for app maintenance during the study duration, as well as the associated data storage costs that are not always accurately estimated. Be prepared to trim down information to be collected.

7 Screenshots of MS Mosaic App

In Figure 3, we present a range of different screenshots from the MS Mosaic App. Figure 4 shows the different surveys that were presented to the participants. We present in Figure 5 a screen showing the symptom survey and how the user could choose a severity in 5-point Likert scale. Figure 6 demonstrates how the data collected from various surveys and functional tests were visualized and shown back to the participants. The MS Mosaic App was also integrated with the Apple health app to obtain passive signal data as shown in Figure 7.

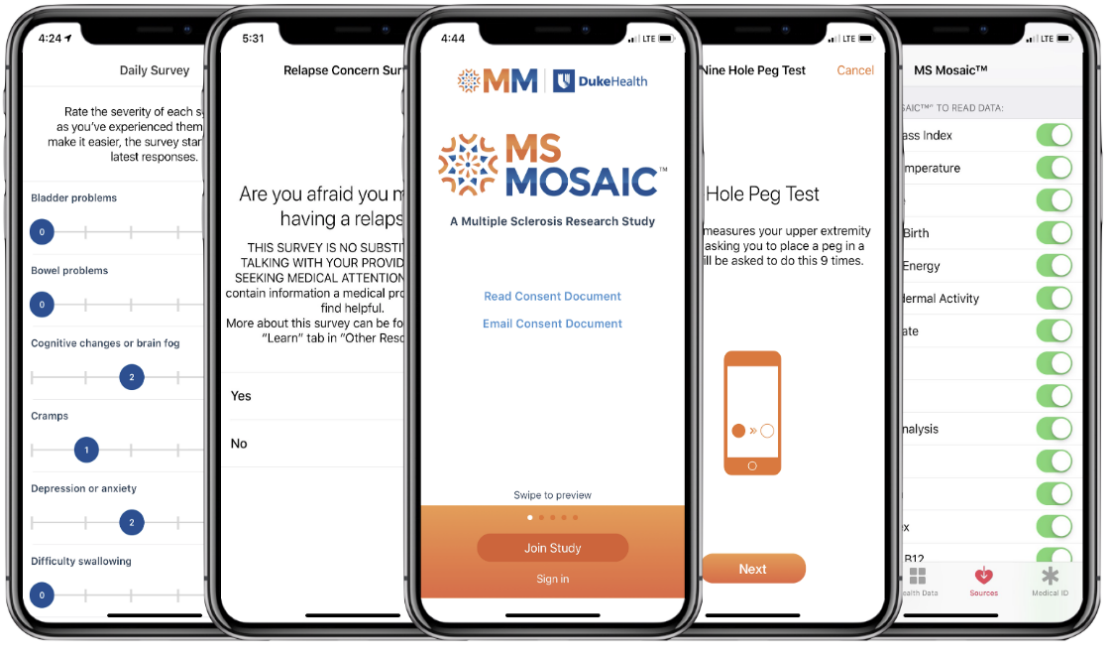

Figure 3: Screenshots from the MS Mosaic app.

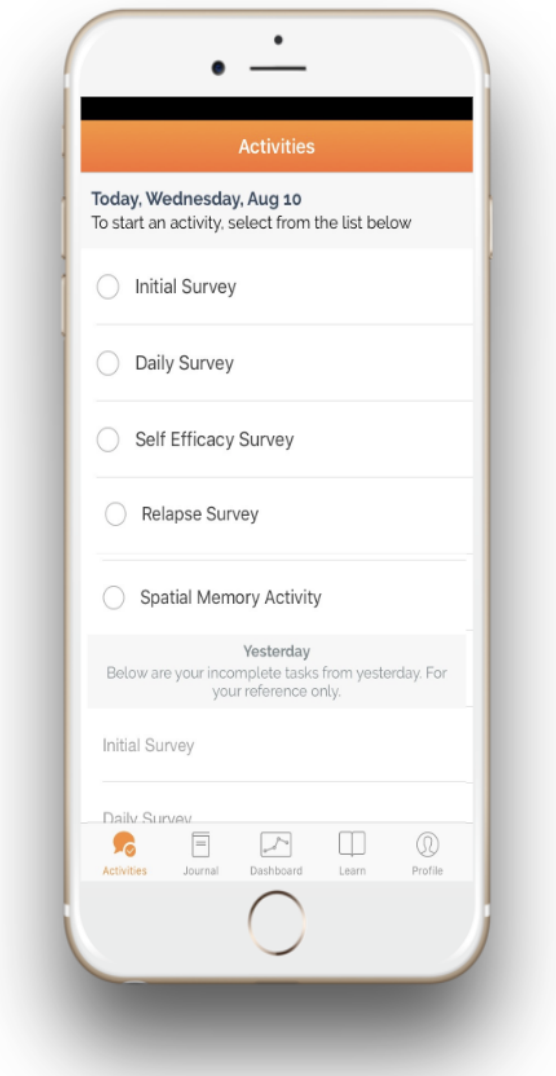

Figure 4: Screen showing the surveys included in the MS Mosaic app.

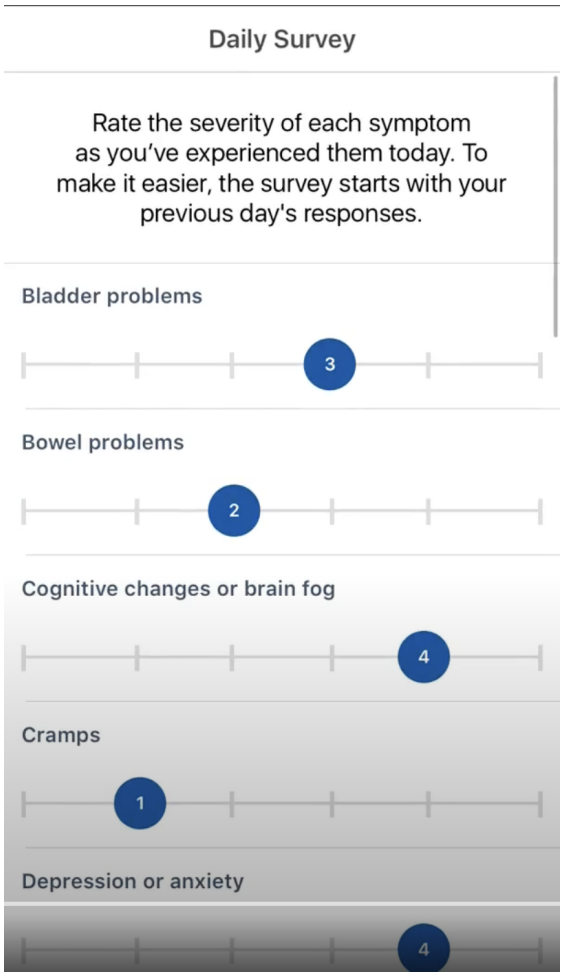

Figure 5: Screen showing the surveys included in the MS Mosaic app.

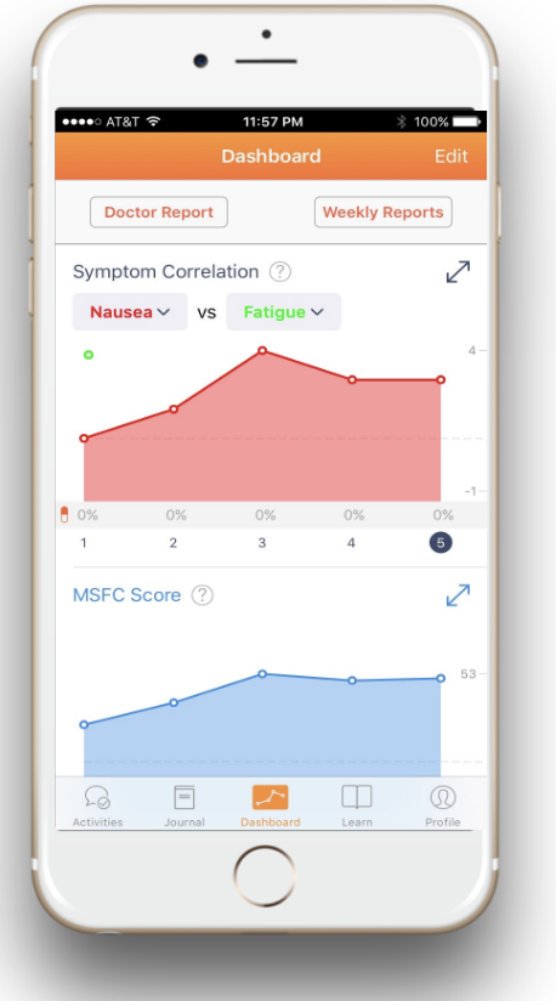

Figure 6: Screen showing visual representation of the symptom data reported by the participants.

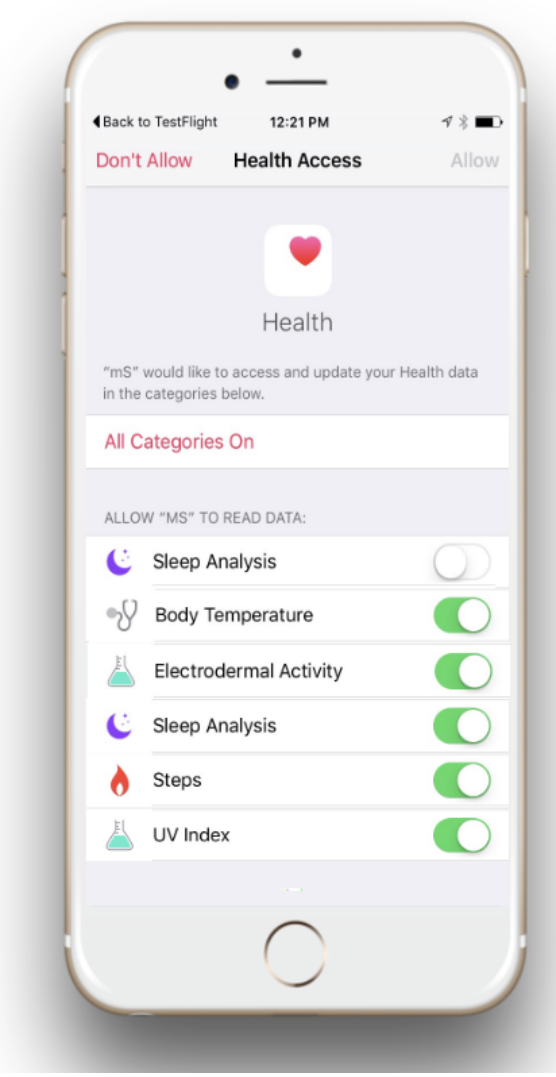

Figure 7: The MS Mosaic app was integrated with Health Kit for passive data collection through wearables.

8 Informed Consent Document

TITLE: MS Mosaic — A Multiple Sclerosis Research Study SPONSOR: Duke University and Duke Neurology

INVESTIGATOR: F. Lee Hartsell,
III MD MPH Duke University Medical Center
DUMC Box 3184
Durham, NC 27710 United States

STUDY-RELATED CONTACT INFORMATION: F. Lee Hartsell, III MD MPH (Principal Investigator)
Katherine Heller, PhD (Co-PI)
919-684-8615 or ms-app@duke.edu

SUMMARY

You are invited to participate in a research study to understand variations in symptoms of multiple sclerosis. This study is designed for persons over 18 years old with or without multiple sclerosis. Your participation in this study is entirely voluntary. To be in a research study you must give your informed consent. The purpose of this form is to help you decide if you want to participate in this study. Please read the information carefully. If you decide to take part in this research study, you will be given a copy of this signed and dated consent form. If you decide to participate, you are free to withdraw your consent, and to discontinue participation at any time. You should not join the research study until all of your questions are answered. Participating in a research study is not the same as receiving medical care. The decision to join or not join the research study will not affect your medical benefits.

WHAT IS A RESEARCH STUDY?

A research study is when scientists try to answer a question about something that we don't know enough about. Participating may not help you or others. People volunteer to be in a research study. The decision about whether or not to take part is totally up to you. You can also agree to take part now and later change your mind. Whatever you decide is okay. It will not affect your ability to get medical care at Duke.

We have provided the contact information for the study's principal investigator above. If you have questions, please ask them. Feel free to ask all the questions you want before you decide. Any new information that develops during this research study that might make you change your mind about participating will be given to you promptly.

Basic information about this study will appear on the website http://www.ClinicalTrials.gov. There are a few reasons for this: the National Institutes of Health (NIH) encourages all researchers to post their research; some medical journals only accept articles if the research was posted on the website; and, for research studies the U.S. Food and Drug Administration (FDA) calls "applicable clinical trials" a description of this clinical trial will be available on http://www.ClinicalTrials.gov, as required by U.S. Law. This Web site will not include informa-

tion that can identify you. At most, the Web site will include a summary of the results. You can search this Web site at any time.

PURPOSE OF THE STUDY

People with multiple sclerosis can have very different and more or less severe symptoms day to day. This affects quality of life and makes managing treatment difficult. We would like to understand the causes of these symptom variations.

New technologies allow people to record and track their health and symptoms in real time. This study will monitor individual's health and symptoms using questionnaires and sensors via a mobile phone application and wearable devices if available.

If you decide to join the study you will need to download the study application on your mobile device. Then periodically we will ask you to answer questions and perform some activities via your mobile phone. These questions may be about your health, exercise, diet, sleep and medicines, in addition to other surveys. The activities will be some brief tasks that you perform while holding your phone like walking, tapping or balancing for a short period of time. In addition, if you are able to sustain moderate physical activity, we may send you motivational prompts to remain active. Your study data will include your responses to surveys and the measurements from the phone itself when you perform an activity. Your data, without your name, will be added to the data of other study participants and analyzed by the study team. Also, if you choose to, your study data (without your contact information) can be made available to other qualified researchers for this and future research. You will have a unique account that you can use to review your own data.

WHO WILL BE MY DOCTOR ON THIS STUDY?

If you decide to participate, Dr. Hartsell is the principal investigator for this study.

SOURCE OF FUNDING FOR THE STUDY

The sponsor is Duke University and Duke Neurology.

How long will I be in the research study?

We anticipate this study will last about six months, however the app can remain on your phone for multiple years, and you can keep using it to track your symptoms and review your data.

How many people will take part in this study?

The number of people expected to take part in this research study is difficult to predict but may be hundreds to tens of thousands of people.

PROCEDURES

What will you be asked to do?

You will first need to download the free study application on your mobile device, review the consent information embedded within the app (identical to what is provided in this consent form), consent to participate in the study, and register an account. Then, periodically we will ask you to answer questions and/or perform activities on your mobile phone. Your study data will include your responses to surveys and activities and some measurements from the phone

itself.

- Register to the study: You will follow the prompts on the app to register an account and confirm your agreement to participate in this study. There will be an electronic consent process explaining the risk s and benefits of using the app. Registration will include entering your name, email address and other general information about yourself to verify your eligibility. You can cancel the registration process at any time.

- Health Surveys: We will periodically ask you to answer questions about yourself, your medical history, and your current health and symptoms to track changes. You may skip any questions that you do not wish to answer blank.

- Activities: We will ask you to perform specific tasks weekly while holding or using your mobile phone and record sensor data directly from your phone. Examples are:
  - to hold your phone and walk 25 steps to assess your walking speed and your gait.
  - to tap on the phone screen in a specific way to test your reaction time and dexterity. Additionally, you may be asked for your permission to include some data from third-party fitness devices (like the Nike FuelBand, Jawbone Up, Withings Activité, or Misfit) if you use one.
  - to repeatedly move a virtual peg across the screen to a particular location
  - to rapidly calculate sums using a series of numbers given by the phone

We will send notices on your phone asking you to complete these activities and surveys. You may choose to act at your convenience, (either then or later) and you may choose to participate in all or only in some parts of the study. Registration may take 20 minutes to complete. The daily surveys should take you no more than 1 minute to complete, while the weekly surveys should take no more than 15 minutes. The prompted tasks should take no more than 5 minutes to complete. You have the right to refuse to answer particular questions or participate in particular aspects of the study.

What we will and will not do with the data?

- Data Processing: We will electronically process your data and separate your account information (name, email, contact information, etc.) and all identifiable information from your study data (your responses to surveys and the measurements from the phone itself when you perform activities). We will combine your de-identified coded study data (without your name) with those of other study participants. The combined data will be transferred electronically to cloud servers hosted by Thread Research and their contracted partners for storage and analysis.

- Use in research: The research team will analyze the combined de-identified data and report findings back to the community through Blog or scientific publications. Your de-identified study data (without your contact information) may also be used by other researchers and for other research purposes if you choose to share them more broadly. The Principal

Investigator and Sponsor will have no oversight on the future use of the de-identified shared study data.

RISKS, DISCOMFORTS, AND INCONVENIENCES There are risks, discomforts, and inconveniences associated with any research study. These deserve careful thought.

- This is not a treatment study and we do not expect any medical side effects from participating.
- Some survey questions may make you feel uncomfortable. Know that the information you provide is entirely up to you and you are free to skip questions that you do not want to answer.
- Other people may glimpse the study notifications and/or reminders on your phone and realize you are enrolled in this study. This can make some people feel self-conscious.
- Be safe – just as you would not text while driving, do not fulfill study tasks activities while driving. Wait until you are in a safe place to perform study-related activities!
- We take great care to protect your information, however there is a slight risk of loss of privacy. This is a low risk because we separate your personal information (information that can directly identify you, such as your name or phone number) from the research data to respect your privacy. However, even with removal of this information, experts in re-identification may be able to reverse our processes and/or attempt to re-identify an individual given enough cross-reference information about him or her.
- Accidental public disclosure may occur due to unintended data breaches including hacking or other activities outside of the procedures authorized by the study. In such a case, your data may be misused or used for unauthorized purposes.
- Data collected in this study will count against your existing mobile data plan. You may configure the application to only use WiFi connections to limit the impact this data collection has on your data plan.
- Participation in this study may involve risks that are not known at this time. You will be told about any new information that might change your decision to be in this study.

POTENTIAL BENEFITS

The goal of this study is to create knowledge, which can benefit us as a society. The benefits are primarily the creation of insights to help current and future patients and their families to better detect, understand and manage their health. We will return the insights learned from analysis of the study data through the study website, blogs and/or research publications, but these insights may not be of direct benefit to you. We cannot, and thus we do not, guarantee or promise that you will personally receive any direct benefits from this study. However you will be able to track your health and export your data at will to share with your medical doctor and anyone you choose.

PAYMENT
You will not be paid for being in this study.

COSTS
There is no cost to you to participate in this study other than to your mobile data plan if applicable.

ALTERNATIVES
Since no medical treatments are provided during this study there are no alternative therapies. The only alternative is to not participate.

AUTHORIZATION TO USE AND DISCLOSE INFORMATION FOR RESEARCH PURPOSES
Because information about you and your health is personal and private, it generally cannot be used in a research study without your written authorization. If you sign this consent form, you will provide that authorization. You do not have to sign this form. But if you do not, you will not be able to participate in this research study.

What personal information will be used or disclosed? Your personal health information that may be used or disclosed in connection with this research study, may include, but is not limited to your body height, weight, gender, age, ethnicity/race, health history, answers to study questions, and health information that may be discernible from your mobile phone's sensors. Your account information, study data and signed consent form may also be looked at and/or copied by designated personnel for regulatory and quality assurance.

Who may use and disclose my data? The study sponsor, investigators, study coordinators and study staff may use and disclose your de-identified study information to do the research described above or as required by law (e.g. to prevent possible injury to yourself or others).

Who May Receive or Use the Information?
The parties listed in the preceding paragraph may disclose your health information as required by law to:

- The US National Institute of health, National Cancer Institute, Office for Human Research Protection, The Food and Drug Administration and other agencies as required,
- Governmental agencies in other countries,
- Duke Institutional Review Board or other Institutional Review Board who watch over the safety, effectiveness and conduct of the research,
- Others, if the law requires

When will my authorization expire?
Your authorization for the use and/or disclosure of your health information will expire if you choose to withdraw from the research study or by December 31, 2060. If you choose to withdraw from the research study, we will stop collecting your study data. At the end of the study period we will stop collecting your data, even if the application remains on your phone and you

keep using it. If you were interested in joining another study afterward, we would ask you to complete another consent, like this one, explaining the risks and benefits of the new study.

CONFIDENTIALITY

Study records that identify you will be kept confidential as required by law. Federal Privacy Regulations provide safeguards for privacy, security, and authorized access. Except as required by law or for your care, you will not be identified by name, social security number, address, telephone number, or any other direct personal identifier. The data collected through the app will be encrypted on the smartphone, transferred electronically and stored securely within a data repository managed by Thread and their contracted partners. Your contact information, including your name and e-mail address will be stored separately from the study data. We will use a random code number instead of your name on all your study data. This code cannot be used to directly identify you. Information about the code will be kept in a secure system.

We will not access your personal contacts, other applications, text or email message content, or websites visited. We will never sell, rent or lease your contact information.

COMPENSATION FOR INJURY

Immediate necessary medical care is available at Duke University Medical Center in the event that you are injured as a result of your participation in this research study. However, there is no commitment by Duke University, Duke University Health System, Inc., or your Duke physicians to provide monetary compensation or free medical care to you in the event of a study-related injury.

For questions about the study or research-related injury, contact Lee Hartsell at 919-684-8615 during regular business hours and by pager at 919-970-6894 after hours and on weekends and holidays.

VOLUNTARY PARTICIPATION AND WITHDRAWAL

Your participation in this study is voluntary. You do not have to sign this consent form. But if you do not, you will not be able to participate in this research study. You may decide not to participate or you may leave the study at any time. Your decision will not result in any penalty or loss of benefits to which you are entitled.

- You should not feel obligated to participate in this study.
- Your questions should be answered clearly and to your satisfaction.
- You have a right to download or transfer a copy of all of your study data.
- By agreeing to participate you do not waive any of your legal rights.

To withdraw from this study please contact the Study Principal Investigator Dr. Lee Hartsell, by email ms-app@duke.edu or call 919-684-8615.

Although you can withdraw from the study at any time, you cannot withdraw the coded study data that have already been distributed. If you withdraw from the study, we will stop collecting new data but the coded data that you have already provided will not be able to be destroyed or

deleted. The Study Principal Investigator or the sponsor may also withdraw you from the study without your consent at any time for any reason, including if it is in your best interest, you do not consent to continue in the study after being told of changes in the research that may affect you, or if the study is cancelled.

USE OF DATA FOR FUTURE RESEARCH

This study gives you the option to share your coded study data more broadly, with other researchers worldwide for use in this research and beyond to benefit future research.

If you choose to share your data broadly, your coded data (without your contact information) will be added to a shared study dataset on servers managed by Thread and their contracted partners. This shared study dataset will be made available to qualified researchers who are registered users of Duke's systems and who have agreed to using the data in an ethical manner, to do no harm and not attempt to re-identify or re-contact you unless you have chosen to allow them to do so. No name or contact information will be included in this shared study dataset. Researchers will have access to the shared study data but will be unable to easily map any particular data to the identities of the participants. The Principal Investigator and Sponsor will have no oversight on the future use of the shared study data by other researchers.

WHOM DO I CALL IF I HAVE QUESTIONS OR PROBLEMS?

For questions about the study or a research-related injury, or if you have problems, concerns, questions or suggestions about the research, contact Dr. Hartsell at 919-684-8615 at any time.

For questions about your rights as a research participant, or to discuss problems, concerns or suggestions related to the research, or to obtain information or offer input about the research, contact the Duke University Health System Institutional Review Board (IRB) Office at (919) 668-5111.

STATEMENT OF CONSENT

"The purpose of this study, procedures to be followed, risks and benefits have been explained to me. I have been allowed to ask questions, and my questions have been answered to my satisfaction. I have been told whom to contact if I have questions, to discuss problems, concerns, or suggestions related to the research, or to obtain information or offer input about the research. I have read this consent form and agree to be in this study, with the understanding that I may withdraw at any time. I have been told that I will be given a signed and dated copy of this consent form."

Signature of Subject

Daten

Time

Signature of Person Obtaining Consent

(Optional):

Date

Time

Signature of Principal Investigator

Date

Time

If applicable, add the following:

Signature of Legal Representative

Date

Time

Relationship to Subject

References

[1] Subhrajit Roy, Diana Mincu, Lev Proleev, Negar Rostamzadeh, Chintan Ghate, Natalie Harris, Christina Chen, Jessica Schrouff, Nenad Tomasev, Fletcher Lee Hartsell, and Katherine Heller. Disability prediction in multiple sclerosis using performance outcome measures and demographic data, 2022. URL https://arxiv.org/abs/2204.03969.