A framework for hybrid machine and human computation for the accurate and scalable analysis of human clinical EEG recordings

1 Aims

Electroencephalography (EEG) is a key tool in the diagnosis of epilepsy and sleep disorders. In current practice, EEG recordings are visually analyzed by specialist technicians and physicians. However, this is a slow process and dependent on the availability of these expert annotators. This limits EEG as a diagnostic medical tool in smaller communities. Even in larger centers, the time consuming nature of human EEG analysis can lead to backlogs and significant delays in diagnosis and treatment. Thus, there is a need for rapid, scalable, cost-efficient, accurate EEG interpretation that is not heavily dependent on the time of highly trained specialists.

These limitations have motivated efforts to develop fully automated algorithms for the interpretation of human EEG signals. Unfortunately, to date, these approaches have had limited success, in part because many aspects of EEG interpretation are fundamentally image classification problems that while straightforward for trained humans, are difficult to fully automate. The field of human computation developed to address such problems where humans may have some advantages over pure automation. In this paradigm, problems are decomposed into a massive number of very simple, carefully designed, human micro-tasks to be carried out by an array of "human processors," whose answers can be combined with automated algorithms to solve the original problem. This maximizes the advantages of both humans and computers in one algorithm, overcoming the limitations of either.

The overall goal of this project is to design a framework for hybrid machine and human computation to achieve accurate and scalable analysis of human clinical EEG recordings in both resource-rich and resource-poor health settings. In collaboration with our knowledge organizations the Bhutan Epilepsy Project, the Epilepsy Program at the Toronto Western Hospital, and the Clinical Neurophysiology Laboratory at Sunnybrook Health Sciences Centre, we propose to achieve the following scientific and translational (i.e., knowledge transfer to hospitals and industry partners) aims:

Scientific Aim 1. Develop a general set of algorithms to decompose EEG analysis into micro-tasks, and integrate the responses of non-expert and expert human processors with automated algorithms to solve EEG-related clinical problems.

Scientific Aim 2. Develop a general framework to compare these algorithms against fully automated approaches and specialist analyses, and an iterative approach to improve these algorithms.

Translational Aim 1. Apply the algorithms to the interpretation of clinical EEGs from Canadian hospitals.

Translational Aim 2. Integrate these algorithms with a EEG smartphone-based recording device developed by the Bhutan Epilepsy Project to allow recording and interpretation of EEGs in remote settings.

The outcomes of this work will include two prototype systems freely available for general use by health care organizations – one for analyzing conventional EEG data from large Canadian hospitals, and one integrated with the smartphone EEG recording system developed by the Bhutan Epilepsy Project for use in rural and remote settings by health care professionals (e.g., nurses) without EEG training. For Canadian hospitals, the benefits will include improved robustness, speed, and efficiency of EEG interpretation, resulting in more timely and accurate diagnosis of epilepsy, sleep disorders and other neurological conditions, as well as more efficient utilization of scarce specialist resources. For more remote regions in Canada and internationally, the integrated smartphone-based version will provide access to accurate, scalable, cost- and time-efficient local EEG diagnosis, with minimal need for outside EEG specialists.

2 Background

2.1 EEG Analysis

Electroencephalography is a key neurological test used in many clinical contexts [18, 20]. Routine 20min EEGs (rEEG) are used in both inpatient and outpatient settings to diagnose epilepsy and other neurological conditions. Longer continuous EEGs (cEEG) recorded 24hr a day are used in inpatient epilepsy monitoring units (EMUs) to characterize seizures in patients with epilepsy, and in intensive care units (ICUs) for real-time detection of seizures and neurological deterioration. In the context of polysomnograms (PSGs), 8hr sleep EEGs (sEEG) are used to quantify sleep architecture and diagnose sleep disorders.

There are four main tasks in EEG interpretation (see Figure 3–8): 1) characterizing dominant frequencies, 2) identifying normal features such as K-complexes, sleep spindles etc, 3) using these features to classify segments of the EEG into normal (W, N1, N2, N3, R) or abnormal (e.g. encephalopathy) states, and 4) identifying discrete pathological features, notably interictal epileptiform discharges (IEDs; brief, often <200ms discharges of characteristic morphology) and seizures (runs of abnormal rhythmic brain activity lasting seconds to minutes which can cause loss of consciousness and other neurological symptoms).

Challenges in EEG interpretation include the need to identify relatively brief and rare features (e.g. a single <200ms IED in a 24-hour recording), the presence of signal noise and artifact that can mask or imitate pathological features, the similarity between normal (e.g. vertex waves) and pathological (e.g. IEDs) features, and the desirability in some contexts (e.g. cEEG in the ICU) for near-real-time identification of abnormalities that may indicate neurological deterioration warranting prompt clinical action.

In current practice, specialized technicians and neurologists visually examine EEGs and report their findings. This has several drawbacks. The dependence on large amounts of specialist time (e.g. interpreting 24 hours of cEEG can take 3 hours of technician and 1 hour of neurologist time) limits EEG analysis to centers with these specialists, limits the number of EEGs performed, drives up costs, can delay diagnosis, and takes specialists way from other clinically vital tasks. Meanwhile, the subjective nature of visual analysis leads to inter-rater variability, making it difficult to compare records interpreted by different physicians. Finally, the "needle in a haystack" nature of EEG analysis makes it prone to lapses in attention.

These limitations have led to a desire for automated approaches to EEG analysis. Seizure detection has attracted particular interest, and many automated algorithms [11, 3, 10, 21] have been developed for this purpose. The complexity in creating automated algorithms comes from multiple factors. First, the mechanisms underlying seizure generation are not well understood, resulting in feature engineering challenges. Second, EEG recordings are made from small electrodes that capture a noisy state of the brain activity at a particular location. The electrode placement may not be optimal, yielding weak signals, and the origin of the seizure may be difficult to determine from the ambiguous data. Third, the relatively low rate of occurrence of seizures and their short duration, compared to the amount of between seizures data, makes the EEG recordings highly unbalanced, posing challenges to most standard machine learning algorithms. An automated seizure detection system must address all these challenges in order to be effective.

Another EEG analysis task that has attracted interest is that of identifying normal features such as sleep spindles, and integrating this information into a segment-by-segment classification of behavioral state. Unfortunately, automated algorithms for the detection of specific normal features (e.g. spindles) perform quite poorly compared to human experts, even when relatively artifact-free recordings are used [17]. Meanwhile, automated approaches to segment-by-segment EEG state classification based on frequency analysis have shown some success when applied to relatively artifact-free datasets from healthy individuals or laboratory animals [8, 1], but have not been tested or proven in populations with a high prevalence of neurological diseases or sleep disorders – the individuals most likely to require EEGs.

2.2 Human Computation

EEG interpretation is an example of a larger class of problems that are trivial for human experts but hard for pure artificial-intelligence (AI) systems. A new field of computer science, *human computation* [7], emerged around the idea of harnessing human intelligence to improve upon automated algorithms. In this paradigm, complex tasks (e.g. image classification) are decomposed into many micro-tasks (e.g. identifying shape, color), each of which may take seconds. Micro-tasks amenable to AI are performed by machines, while those that are not are performed by humans. The output of various micro-tasks is then aggregated to solve the original problem. Two key insights underlie human computation. First, by combining AI algorithms with human input, hybrid systems capture the strengths of both while avoiding the weaknesses of either. Second, by efficiently integrating the input of many non-experts and augmenting this with AI algorithms, the accuracy of groups of non-experts can approach or exceed that of experts.

Human computation has been used to solve a host of scientific problems that had proven intractable to pure AI methods, including protein folding [5], object recognition [12, 19], and machine translation [13]. Examples of human computation systems include crowdsourcing marketplaces (e.g., Amazon Mechanical Turk [9]) where workers perform micro-tasks for payment, games with a purpose [15, 6, 14] that generate useful data through gameplay, identity verification systems, e.g., reCAPTCHA [16], which digitize books through billions of users performing computation (e.g. transcribing words) for access to online content. Human computation approaches have been successfully applied to the analysis of some forms of clinical time series data. Zhu et al, in collaboration with China Mobile, have implemented a human computing platform for electrocardiogram (ECG) interpretation for a leading cardiac hospital in Beijing [23]. By efficiently decomposing the task of ECG interpretation and integrating the input of multiple experts using a Bayesian framework, substanital improvement in accuracy is achieved [22].

Most recently, in a paper in *Nature Medicine*, Warby et al [17] applied human computation to the detection of sleep spindles in human EEGs. In their system, the aggregate performance of groups of non-experts exceeded that of even the best automated algorithms, and approached that of trained experts. Moreover, they showed that aggregating the input of groups of experts resulted in performance superior to any single expert. While this work concerns only sleep spindles, it demonstrates the potential superiority of human computation approaches compared to fully automated approaches for EEG feature detection. It also suggests that a carefully designed system combining human computation and automated algorithms for the detection of EEG features may achieve the accuracy of a trained expert, with minimal expert input.

3 Progress to Date

A core component of any hybrid human-machine computing system is a platform to decompose complex tasks into micro-tasks, dynamically delegate these tasks to non-experts, experts, and automated algorithms, and reintegrate their results. PI Edith Law has developed such a platform. The Curio project (Figure 2) is general platform designed to allow scientists to apply human computation approaches to data processing tasks not easily amenable to pure AI approaches. It includes mechanisms for task decomposition, task delegation and result visualization. In collaboration with co-I Andrew Lim, Dr. Law has begun to adapt the interfaces and algorithms on the Curio platform to handle clinical EEG data.

Co-I Joelle Pineau has previously developed a machine learning toolkit for automating the various phases of building a personalized epileptic seizure detection algorithm. It provides a pool of features and parameterizations, a feature selection module that tailors algorithms to specific patients, high-level machine learning classifiers for detecting seizure events, and a new dataset containing EEG recordings from rats with epilepsy. Our research will further develop ideas on how these algorithms can be combined with human annotations to achieve better performance.

4 Research Plan

4.1 Sources of Clinical Data

Our goal is to design a framework for hybrid machine and human computation for the analysis of human clinical EEGs. Toward this end, we will leverage clinical EEG data from our knowledge organizations:

1) The Toronto Western Hospital / University Health Network (UHN) epilepsy program is one of the largest in Ontario, with over 1000 patient-visits per year. A key component is the clinical neurophysiology laboratory and associated inpatient epilepsy monitoring unit. The laboratory performs over 12,500 rEEGs and obtains over 84,000 hours of cEEG recording per year, and has an archived database of over 30,500 rEEGs, and over 680,000 hours of cEEG, all with expert annotation. This will provide fully annotated normal and abnormal rEEG and cEEG recordings from across the clinical spectrum.

2) The Sunnybrook Clinical Neurophysiology Laboratory is a major clinical neurophysiology site located at Sunnybrook Health Sciences Centre, a leading Canadian teaching hospital. The laboratory performs over 1300 rEEGs and nearly 1500 sEEGs per year, and has a database of over 6500 sEEGs and 4300 rEEGs. This will give us access to normal and abnormal sEEGs from across the clinical spectrum, as well as an independent set of rEEGs

3) The Bhutan Epilepsy Project is a non-governmental multi-national organization that is developing a smartphone-based portable EEG recording device to enable low-cost EEG recordings in remote regions with minimal specialist expertise. The Project is testing prototype devices in Bhutan and is amassing a collection of 200 recordings, to compare with 200 recordings from a conventional EEG device. We will have access to both sets of recordings, enabling us to compare the performance of our algorithms across recording platforms, and to adapt our framework to more resource-limited settings.

4.2 Aim 1: Methodology

Our first aim is to develop a general set of algorithms to decompose clinical EEG analysis into micro-tasks, and coordinate input from experts, non-experts and automated algorithms. We will develop a **first-pass filter approach** whereby each record is analyzed first by automated algorithms. Selected epochs will be passed to non-experts for additional semantic feature annotation. We will then combine the automatically extracted features and features identified by non-experts to generate a classifier. Uncertainty scores from this classifier will determine specific epochs or features to send to experts for further processing.

We aim to develop machine learning algorithms for combining expert labels, features detected by non-experts and automated algorithms, with a particular interest towards algorithms that can provide confidence estimates. Ensemble methods [4], for example, compute confidence based on agreement between multiple classifiers trained on the same data. Co-training [2], on the other hand, derives confidence from the same classifier trained on two different sets of features which are assumed to be conditionally independent given the class label, such as automatically extracted EEG features versus semantic EEG features (e.g. presence and absence of sleep spindles) scored by human annotators. This semi-supervised technique has the added advantage of requiring only a small number of labeled examples from experts to begin with; the algorithm will iteratively add more labeled examples if the two classifiers agree.

We describe here our proposed methodology for integrating automated algorithms, non-expert input, and expert input to solve each of the key tasks in EEG interpretation described in section 2.1.

Frequency Analysis. A key first step in EEG interpretation is the characterization of segment-by-segment dominant frequencies, which can influence how the discrete features identified in the other tasks are interpreted. We propose to fully automate this using standard signal analysis techniques e.g. Fourier analysis.

Identification of Normal Features. The normal EEG is characterized not only by background rhythmic activity, but also by discrete morphological features that define the normal five behavioral states (W, N1, N2, N3, and R) and whose absence may be indicative of an abnormal state (e.g. encephalopathy). We will apply a first pass filter approach here. Initially, machine algorithms will be used to identify the presence of various normal features in 30s epochs. Because these features are common, these algorithms will be biased toward specificity, rather than sensitivity. Where features are detected with high confidence, no further human input will be needed. Epochs with features detected at lower confidence, or epochs with no features detected, will be sent to non-experts for additional feature detection, based on characteristics which will be selected in consultation with Co-I Andrew Lim and knowledge user Brian Murray (both neurologists with board certification in sleep EEG interpretation). Others [17] have shown that, in principle, many such features can be confidently identified by non-experts.

Epoch-By-Epoch Classification of Behavioral State. Automated algorithms will combine information from the background frequency analyses, automated feature detection, and non-expert feature detection, to arrive at a classification of the behavioral state for each non-overlapping 30s epoch of the EEG. For most epochs, these features will be concordant and a classification will be made without expert intervention. For ambiguous epochs, experts will be provided with a visualization of a combination of automatic and non-expert annotations along with the uncertainty of their labels, to arrive at a final classification.

Identification of Abnormal Features. The next task is identification of abnormal discrete features, especially IEDs and seizures, as described in 2.1. As above, we will use a first-pass filter approach. Automated algorithms developed by Co-I Joelle Pineau, as described in section 3 above, will first be deployed to prune the record. These will be biased toward high sensitivity to ensure that no features are missed. Then, the remaining portions of the record will be sent to available non-experts (e.g. nursing staff) to evaluate for specific characteristics, which will be selected in consultation with Co-I Andrew Lim and knowledge user Richard Wennberg (both neurologists with board certification in EEG interpretation). These characteristics may include, as an example, the “sharpness” of a discharge or the extent of “evolution” of a run of rhythmic activity identified by an automated algorithm. The input from non-experts and the automated algorithms will then be integrated to create a confidence score for each feature. Features identified with high confidence and concordance may be classified at this stage. Features that are identified with lower confidence or some discordance will then be sent to experts for disambiguation. The benefit of inserting a layer of non-experts between the algorithms and the experts is that many algorithm misclassifications may be easily *corrected* by available non-expert readers without expert intervention. This is particularly important because in many situations (e.g. overnight care, or in smaller communities) non-experts such as nurses may be available, but EEG trained neurologists may not.

Generation of Final Narrative Report. The final output of our framework will be an automatically generated table, based on automated algorithms and non-expert input, with or without additional expert input, indicating the presence/duration of abnormalities or behavioral state, and the presence/timing of discrete abnormalities such as seizures and interictal discharges. This narrative report, along with a fully annotated record, will be provided to the expert (if available) for further analysis and reporting.

4.3 Aim 2: Methodology

The first-pass filter approach developed in Aim 1 will generate an initial set of algorithms and interfaces that will enable a combination of automated algorithms and non-experts to handle a large proportion of most records, routing only ambiguous pages and features to experts. Aim 2 is to develop and validate an **iterative refinement process** to improve these algorithms over time, and increase the proportion of the overall workload that can be handled by the combination of automated algorithms and non-experts.

First, this approach will include algorithms and interfaces to iteratively evaluate and improve the performance of non-experts, experts and automated algorithms. For example, when a non-expert differs from the group consensus or from an expert, the system will present corrective feedback to facilitate improvement. Using the same ground truth data, the accuracy of each non-expert can be re-assessed and recorded, allowing the system to personalize their training. Likewise, experts can improve over time as they are provided with more robust visual aids displaying the crowdsourced features, and the automated algorithms can improve over time due to the addition of training data from experts or the non-expert crowd. Second, a similar iterative refinement approach will be applied to the framework as a whole. As the reliability of non-experts and automated algorithms improves, other algorithms will monitor and re-weigh these different sources of contributions, slowly moving toward a decreasing reliance on human expert input.

To validate the iterative refinement approach, we will engage a group of non-experts and experts in a month-long experiment to evaluate the accuracy and efficiency of each of the components over time, and how the adaptive combination of these components impacts the accuracy of the overall classification and the distribution of the workload.

4.4 Translational Aim 1: Methodology

Translational Aim 1 is to apply our framework to the interpretation of a wide variety of clinical EEGs (rEEG, cEEG, sEEGs) from Canadian hospitals. We will develop an interface between the EEG recording software and our system to allow recordings from consecutive consenting patients (stripped of all personal health information) to be uploaded to a central server in real-time. There, the interpretation of the record will be parsed into micro tasks as described in Aims 1 and 2 above, and distributed to automated algorithms, as well as to non-experts and experts via a web application which can be accessed on participants' own time and from their own locations. For experts, we will use the physicians and technicians at the Toronto Western Hospital Epilepsy Program and the Sunnybrook Health Sciences Centre Clinical Neurophysiology lab, as well as co-I Andrew Lim and a technician from his laboratory, who will provide an independent source of expert annotations. For non-experts, we will use two sources: 1) paid crowdworkers through Amazon Mechanical Turk (as used by [17] in their study of sleep spindle detection) or paid CS undergraduates at the University of Waterloo. All records will also be annotated by experts at the recording site as per the usual clinical workflow to generate gold standard annotations. To ensure that we capture an unbiased representation of EEGs at our knowledge user sites, we will study consecutive recordings, without pre-selecting based on neurological diagnosis, patient features, or recording equipment.

As described in section 5.2, this process will begin with an in-depth needs assessment during which the computer science (CS) team from Dr. Law's lab will embed themselves within the clinical teams at TWH and Sunnybrook to observe the current process of EEG interpretation, identify gaps, and identify key attributes of successful solutions. We will also conduct formal focus groups to allow technicians and neurologists to share perceived problems and solutions with the CS team. Then, as the framework and interfaces are being developed, there will be monthly site visits, during which the technicians and neurologists will use and provide feedback on the latest iterations of the prototype, to ensure a seamless interface between the end users, the clinical workflows, and our system.

To evaluate our approach, we will compare it to standard expert visual analysis. Efficiency criteria will include person-hours of human input needed per recording and at each level of training (non-expert vs. technician vs. neurologist). We will also assess total time to interpretation. For accuracy, we will assess concordance with the expert read on a number of levels - e.g. on the whole record level (e.g. seizure or not anywhere in the recording), on the page-by-page level (e.g. proportion of pages correctly classified as epileptiform or not) or on a feature-by-feature level (e.g. detection of individual IEDs or seizures).

4.5 Translational Aim 2: Methodology

The Bhutan Epilepsy Project (`www.bhutanbrain.com`) is a new multi-national venture, funded in part by the Government of Canada through the *Grand Challenges Canada* program, developing and evaluating a portable EEG device to allow EEG recordings in rural and remote settings without specialized technicians or neurologists. The focus is initially on Bhutan, as an archetypical under-resourced health care setting, but once validated the developed platform may be applicable to rural/remote settings elsewhere including in geographically dispersed industrialized countries such as Canada. Several prototypes (as seen in Figure 1) have been produced and are being tested in Bhutan. They use an open-source software design, and are compatible with existing mobile devices. The devices cost 250-300 dollars per unit.

The existing prototype device facilitates one aspect of EEG-based diagnostic services in remote settings - obtaining the recordings. However, the system is still reliant on transmitting the recorded data to outside specialists for interpretation. We propose to adapt the human computation framework developed in aims 1 and 2 to a mobile computing setting centered on the smartphone-based recording device developed by the Bhutan Epilepsy Project. This will allow not only recording, but also interpretation of EEGs to occur using available local health care workers, minimizing the need for outside experts. The proposed workflow would depend on the availability of a local internet connection. Where a connection is available, the EEG would be captured using the smartphone-based device and transmitted to our central server in real-time with task decomposition and delegation proceeding as described in section 4.2 above, with local health workers constituting a source of non-experts. Where a connection is not available, algorithms on the recording smartphone itself would begin the process of task decomposition and delegation. Micro-tasks delegated to non-experts could be assigned to local personnel (e.g. local health care workers) for completion right on the recording smartphone, or nearby smartphones communicating wirelessly with the recording smartphone. The input of the local non-experts and algorithms running locally would then be integrated, and the goal is that a substantial proportion of most records could be correctly interpreted using local resources alone. Rare ambiguous features would be flagged. These annotations could be transmitted from the smartphone to remote experts once the phone could be brought within range of an internet connection to complete the interpretative process. Finally, we will apply an iterative refinement process described in section 4.2 to the smartphone based platform in remote settings in the same way as for the hospital based setting. In addition to improving our algorithms, this would have the additional advantage of iteratively improving the performance of local health care workers at EEG interpretation.

Overall, we anticipate several potential advantages of our approach compared to the current reliance on outside experts to read EEGs recorded in remote communities. These include 1) development of the capacity to interpret many if not most EEGs locally using a combination of automated algorithms and local health care workers, 2) gradual improvement of the EEG interpreting capabilities of local workers through the iterative refinement process described in Aim 2, 3) minimization of the need for expert input, meaning that one expert could potentially support EEG interpretation in many communities, 4) allowing for "micro-volunteering" where many experts can contribute meaningfully to EEG interpretation in remote communities one micro task at a time.

4.6 Innovative Aspects

Several key innovative aspects separate this proposal from past attempts to automate EEG interpretation:

Hybrid Human-Machine Computation. To our knowledge, this will be the first application of cutting edge hybrid algorithms to the problem of EEG interpretation. This promises to capture many of the advantages of full automation (speed, efficiency, consistency) while overcoming its disadvantages (difficulty with ambiguous cases and more subtle pattern recognition tasks).

Task Decomposition. Careful task decomposition carries many advantages, even if all key steps are performed by human processors without automated algorithms. These include the capacity for massive parallel processing resulting in more rapid turnaround, and the ability to engage large communities of potential contributors while requiring minimal input from each individual.

Mobile Computing Environment. Previous attempts at automated EEG interpretation have been targeted toward hospital-based EEG labs, making them poorly suited to resource-poor settings, such as in remote/rural regions of Canada. While the Bhutan Epilepsy Project has developed the hardware to enable EEG recording in remote settings, our work tackles the software and algorithmic aspects needed for a truly scalable and portable solution to the problem of EEG interpretation in such settings.

4.7 Potential Difficulties and Feasibility

Several key attributes enhance the feasibility of this proposal:

Leveraging Existing Datasets and Frameworks. Through our partners, we have access to a truly vast database, ensuring that algorithm development will not be delayed by slow prospective data acquisition. Moreover, development of web-based interfaces on Curio for crowdsourcing time series data annotation is largely complete, allowing us to focus on algorithm development and integration with clinical workflows.

Existing Evidence of Non-Expert Engagement. There already exist pilot projects for training nurses to watch cEEG recordings, and have them call/page a specialist physician when they see something they think is a seizure. The Bhutan Epilepsy Project itself has seen overwhelming support from the local hospital teams, who are not trained in EEG analysis, but nonetheless are willing to help.

Accuracy of Non-Experts and Automated Algorithms. The reliance on non-experts and automated algorithms may introduce bias, or result in missed features. We will address this by continuously comparing the performance of algorithms, non-experts and experts performing the same tasks, and assigning to experts tasks that are consistently poorly performed by non-experts or algorithms, and adpatively flagging or down-weighing poorly-performing non-experts or algorithms.

4.8 Description of Roles

Investigators. As the lead PI and a world expert in human computation, Edith Law will have the overall responsibility of the project and will lead a team of 2 PhD students and 1 MSc student to adapt the time series component of the Curio platform for clinical purposes. She will develop the human computation techniques, including algorithms for routing tasks and aggregating responses, as well as interfaces for non-expert annotations and training. As a board certified neurologist with specialty certification in interpreting EEGs and PSGs, co-I Andrew Lim will provide detailed input into the design of the task decomposition and integration frameworks. He and his technician will coordinate data transfer between the various study sites, and provide an indpendent source of expert annotations. As an expert in machine learning, Joelle Pineau will be responsible for the development and adaptation of algorithms for seizure detection. She will supervise 1 Ph.D. student (50% time) to extend her existing seizure detection toolbox for this project, and provide support for integration with Curio. As a senior researcher, she will also provide guidance on managing the project. As a neurologist and epidemiologist with expertise in designing and evaluating systems of epilepsy care, co-I Jorge Burneo will play a key role in guiding integration of our technology in the context of epilepsy care, and in evaluating the effectiveness and usability of our tools.

Knowledge Users. Brian Murray, Richard Wennberg, and Farrah Mateen will play key roles in the integrated KT process (see section 5.2) on behalf of the Sunnybrook Clinical Neurophysiology Laboratory, TWH Epilepsy Program, and Bhutan Epilepsy Project, respectively.

5 Impact

5.1 Impact on the Health and Economic Well-Being of Canadians

Approximately 250,000 Canadians, suffer from epilepsy [18]. Meanwhile, the prevalence of sleep apnea, the most common sleep disorder, has been estimated at $\sim$2-4% in adults [20]. EEG is key to the diagnosis of epilepsy, sleep disorders, and other neurological diseases. A large center such as the Toronto Western Hospital can perform over 12,500 EEGs annually. EEG and other tests for epilepsy are estimated to account for $> 21\%$ of the direct costs of caring for epilepsy patients [18]. Visual interpretation of EEG by trained specialists has the drawbacks of being slow, costly, and inefficient. Furthermore, these drawbacks make it difficult to provide EEG-based diagnostics in smaller communities in Canada and internationally.

Our project will address these needs by producing two prototype systems - one for conventional EEG data from large Canadian hospitals, and one integrated with the smartphone EEG recording system developed by the Bhutan Epilepsy Project for use in rural and remote settings by healthcare professionals without EEG training. For **large Canadian hospitals**, there will be several benefits: 1) Timeliness. The use of automated algorithms to prune EEG records will reduce the human input needed to review each record. Meanwhile, careful task decomposition may allow much of the required human input to be provided by available non-EEG trained health care workers. Specialists may be called upon to provide input only on ambiguous cases, thus allowing a large proportion of records to be screened in near real-time without specialist input. 2) Accuracy. By decreasing the hours of recording requiring human review, and allowing for parallel review of ambiguous features by multiple experts, the proposed system will minimize errors due to lapses in vigilance, and improve accuracy especially in ambiguous cases. 3) Cost Efficiency. Specialist labour is the main cost associated with EEGs. By decreasing the need for this, our system should decrease EEG costs. 4) Capacity. By decreasing the amount of specialist labour needed to interpret each EEG, a single specialist will be able to review a much larger number of EEGs. For **smaller Canadian communities** the overriding benefit with be the capacity to obtain EEG recordings locally without transferring patients or relying on outside experts, reducing costs, and allowing more rapid diagnosis and treatment.

5.2 Integrated Knowledge Translation Plan

The UHN/TWH Epilepsy Program, the Sunnybrook Clinical Neurophysiology Laboratory, and the Bhutan Epilepsy Project will be deeply involved in every stage of the knowledge translation process.

Needs Assessment. One of the first activities will be for the computer science (CS) team from Dr. Law's lab to embed for several days within the clinical services of the UHN/TWH Epilepsy Program and the Sunnybrook Clinical Neurophysiology Laboratory. In addition to observing the process of recording and interpreting EEGs, hence gaining familiarity with the nature of the data and the challenges of EEG interpretation, they will also follow the neurologists as they treat patients. This will allow the CS team to appreciate the potential real-world clinical impact of faster, more accurate, and more efficient EEG diagnosis. We will also facilitate a series of structured "focus group" sessions between the CS team and the TWH and Sunnybrook EEG staff to more formally assess gaps in the current model of EEG interpretation.

Iterative Development of Technology. Our partners will be deeply integrated into the development process. They will host monthly visits during which they will be able to test and provide feedback on the prototype. The CS team will formally observe the technicians and neurologists interacting with the software, and iterate the interface. The TWH and Sunnybrook clinical teams will play a key role in ensuring that our technology is seamlessly integrated into the standard clinical workflow. A similar process of monthly meetings by Skype will occur with the Bhutan Epilepsy Project's front-line workers.

Technology Evaluation. Our partners will play a key role in evaluating the technology. Working closely with Dr. Burneo, the clinical teams at TWH and Sunnybrook will develop and measure clinically relevant benchmarks for accuracy, efficiency, and acceptability while comparing our prototype system with standard visual analysis. Meanwhile, under the auspices of the Bhutan Epilepsy Project, we will formally test the smartphone prototype in the field in Bhutan, evaluating not only accuracy and efficiency, but also acceptability to health care professionals in a resource-limited setting. Drs. Pineau and Law will disseminate these results to the CS community through conference presentations and publications in CS journals, while Drs. Burneo, Lim, Wennberg, Murray, and Mateen will do the same in relevant neurology conferences and publications.

Post-Development Dissemination. Our knowledge users are key to our post-development dissemination plan. The Sunnybrook and TWH programs are located at key academic hospitals in Ontario with a tradition of IT leadership. In addition to integrating our technology into their own programs, serving as an example for other centers, they play a key role in training the next generation of technicians and neurologists, who can serve as ambassadors for the technology at their future employers. Having demonstrated the feasibility of our prototype technology in real world settings in Canadian hospitals and in the field in Bhutan, we anticipate the next stage will be to work with colleagues in health care informatics and software engineering to develop a privacy compliant version of the prototype stably and well integrated into widely used health care informatics systems. While this is being done, our knowledge users will continue to play an important role in building clinical acceptance for the technology. Drs. Wennberg and Burneo are deeply involved with the Canadian League Against Epilepsy as past-president and executive chair for education and will assist with disseminating the technology to key epilepsy professionals through this forum. Dr. Murray is chair of the neurology specialty committee of the Royal College of Physicians and Surgeons of Canada – the body that oversees the training of neurologists in Canada – and is also an executive member of the Canadian Sleep Society, and will help to disseminate our technology through these organizations.

5.3 Training of Highly Qualified Personnel

Training of HQPs is a key component of this project. We will train 2.5 CS PhD students and 1 MSc student in Waterloo and McGill. As described above, these students will have deep and ongoing interactions with the clinical teams at Sunnybrook, TWH, and with the BEP. There will be monthly visits to the Sunnybrook and TWH sites, during which they will embed within the clinical teams, both to formally evaluate clinical needs and iteratively evaluate technology prototypes, as well as to gain a broader appreciation of the impact of potential solutions in the care of patients. They will also have the opportunity to do field work in Bhutan to formally evaluate our technology, and to gain a better appreciation of the unique challenges of providing healthcare in resource-poor settings.

In addition to formally training CS PhD students, this project will also provide valuable training to the next generation of technicians, physicians, and neurologists. The Sunnybrook and TWH labs are major training sites for EEG/PSG technicians and neurologists. These trainees will work collaboratively with cutting-edge researchers in machine learning and human computing, and will not only become familiar with the technology and its capabilities, but also begin to think about how this technology can be applied outside of the EEG setting to other areas of health care.