Author/s:
Payne, DE;Dell, KL;Karoly, PJ;Kremen, V;Gerla, V;Kuhlmann, L;Worrell, GA;Cook, MJ;Grayden, DB;Freestone, DR

Title:
Identifying seizure risk factors: A comparison of sleep, weather, and temporal features using a Bayesian forecast

Date:
2021-02-01

Citation:
Payne, D. E., Dell, K. L., Karoly, P. J., Kremen, V., Gerla, V., Kuhlmann, L., Worrell, G. A., Cook, M. J., Grayden, D. B. & Freestone, D. R. (2021). Identifying seizure risk factors: A comparison of sleep, weather, and temporal features using a Bayesian forecast. Epilepsia, 62 (2), pp.371-382. https://doi.org/10.1111/epi.16785.

MR. DANIEL ERIC PAYNE (Orcid ID : 0000-0002-8608-0253)

DR. PHILIPPA J KAROLY (Orcid ID : 0000-0002-9879-5854)

DR. LEVIN KUHLMANN (Orcid ID : 0000-0002-5108-6348)

PROF. DAVID B GRAYDEN (Orcid ID : 0000-0002-5497-7234)

Article type : Full length original research paper

Full title: Identifying seizure risk factors: A comparison of sleep, weather and temporal features using a Bayesian forecast

Authors: Daniel E. Payne[1,2], Katrina L. Dell[2], Phillipa J. Karoly[1,3], Vaclav Kremen[4,5], Vaclav Gerla[5], Levin Kuhlmann[2,6], Gregory A. Worrell[4], Mark J. Cook[2,3], David B. Grayden[1,2], Dean R. Freestone[2]

Affiliations: [1]Department of Biomedical Engineering, The University of Melbourne, Melbourne, VIC, AUS; [2]Department of Medicine, St Vincent's Hospital, The University of Melbourne, Melbourne, VIC, AUS; [3]Graeme Clark Institute, The University of Melbourne, Melbourne, VIC, AUS; [4]Department of Neurology, Mayo Clinic, Rochester, MN, USA, [5]Czech Institute of Informatics, Robotics, and Cybernetics, Czech Technical University in Prague, Prague, CZ; [6]Department of Data Science and AI, Faculty of IT, Monash University, Clayton VIC, Australia.

Corresponding Author: Daniel Payne

d.payne@student.unimelb.edu.au

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/EPI.16785

(+61) 425783098, Biomedical Engineering

The University of Melbourne, Parkville, VIC 3010, Australia,

Keywords: Seizure, Forecasting, Sleep, Weather, Circadian

Number of Text Pages: 20

Number of Words: 3992

Number of References: 49

Number of Figures: 5

Number of Tables: 1

ORCID numbers:

Daniel E. Payne, 0000-0002-8608-0253

Katrina L. Dell, 0000-0002-3552-6411

Phillipa J. Karoly, 0000-0002-9879-5854

Vaclav Kremen, 0000-0001-9844-7617

Vaclav Gerla, 0000-0002-6616-2112

Levin Kuhlmann, 0000-0002-5108-6348

Gregory A. Worrell, 0000-0003-2916-0553

Mark J. Cook, 0000-0002-8875-4135

David B. Grayden, 0000-0002-5497-7234

Dean R. Freestone, 0000-0003-2411-5358

Summary

Objective: Most seizure forecasting algorithms have relied on features specific to electroencephalogram recordings. Environmental and physiological factors, such as weather, or sleep, have long been suspected to affect brain activity and seizure occurrence but have not been fully explored as prior information for seizure forecasts in a patient-specific analysis. The study aimed to quantify whether sleep, weather and temporal factors (time of day, day of week

and lunar phase) can provide predictive prior probabilities that may be used to improve seizure forecasts.

Methods: This study performed post-hoc analysis on data from eight patients with a total of 12.2 years of continuous intracranial electroencephalogram recordings (average 1.5 years, range 1.0-2.1 years), originally collected in a prospective trial. Patients also had sleep scoring and location-specific weather data. Histograms of future seizure likelihood were generated for each feature. The predictive utility of individual features was measured using a Bayesian approach to combine different features into an overall forecast of seizure likelihood. Performance of different feature combinations was compared using the area under the receiver operating curve. Performance evaluation was pseudo-prospective.

Results: For the eight patients studied, seizures could be predicted above chance accuracy using sleep (five patients), weather (two patients) and temporal features (six patients). Forecasts using combined features performed significantly better than chance in six patients. For four of these patients, combined forecasts outperformed any individual feature.

Significance: Environmental and physiological data, including sleep, weather and temporal features, provide significant predictive information of upcoming seizures. Although forecasts did not perform as well as algorithms that use invasive intracranial electroencephalography, the results were significantly above chance. Complementary signal features derived from an individual's historic seizure records may provide useful prior information to augment traditional seizure detection or forecasting algorithms. Importantly, many predictive features used in this study can be measured non-invasively.

Keywords: Seizure, Epilepsy, Sleep, Weather, Circadian, Rhythms

Key Points Box

- For some patients, weather, sleep and temporal features contain significant weak to strong predictive information about upcoming seizures.
- Seizure likelihood distributions varied between individuals, and the best performing feature combinations were highly patient-specific.
- Features combined using a Bayesian approach provide forecasts that outperformed single features

Introduction

One of the most debilitating aspects of living with epilepsy is the unpredictability of seizures[1,2]. Seizure forecasting may reduce the burden of uncertainty and improve quality of life for people with epilepsy[3]. With accurate seizure forecasting, people with epilepsy could take precautions when seizure risk is high and participate in a wider range of activities when seizure risk is low. Previous work has shown that seizure prediction is possible[4–6] but that seizure forecasting algorithms need to be patient-specific[7]. Most seizure prediction studies used very few seizures per patient[8–10], making it difficult to develop patient-specific algorithms. On the other hand, generic algorithms fail to generalise to new data, making them unreliable for individual patients[11].

Seizure prediction generally uses intracranial EEG (iEEG), using features such as power spectrum[12–14], synchronicity[15–17], etc. Such features are typically fed into machine learning algorithms with varying degrees of complexity[18,19]. Results can be made more robust when multiple algorithms and features are combined[5]. The reliance upon intracranial EEG requires invasive surgery, which may deter many potential users of a forecasting device.

People with epilepsy often report factors that correlate with seizure occurrence, such as sleep deprivation, stress and weather[20]. Numerous measurements that may not require iEEG have been proposed for use in seizure forecasting, including cortisol levels, heart rate, weather, sleep, etc.[2] Forecasting using these factors is gaining interest. However, studies investigating their utility remain limited, partly due to lack of data. Retrospective analysis of existing data is limited by the number of seizures per patient and the types of measurements recorded. Some features can be inferred retrospectively from existing iEEG (sleep patterns, circadian rhythms) or other sources, such as historical weather data.

This study investigates the utility of different environmental and physiological factors for seizure forecasting: sleep, temporal features (clock and calendar time) and weather. Continuous long-term EEG recordings[4] were used to establish the relationship between these factors and seizure occurrence at a patient-specific level, enabling seizure forecasting to be compared for different feature types. Recent studies have shown the forecasting potential of sleep, weather and temporal factors. Patient-specific circadian and multiday cycles of seizure occurrence are well established[21–23]. Temperature, humidity and pressure may also correlate with seizure occurrence[24], although no patient-specific analysis has yet been performed. Sleep deprivation has long been considered to increase seizure frequency[25,26] although this has recently been

contradicted[27] and the subtleties of the relationship between sleep and seizures are not fully understood.

Given that forecasting performance can be improved with multivariate models, it is well worth investigating the predictive power of combining information from complementary signals. To combine features into a single forecast, machine learning algorithms are commonly used. These algorithms often have many parameters, making them effective at finding patterns in multivariate data. However, this approach requires a large volume of data, meaning hundreds of seizures may be needed for an effective forecaster to be developed. In contrast, a Bayesian approach that combines forecasts from several factors is simple and flexible[28]. Inputs can be changed without the need to retrain the whole algorithm. The Bayesian approach may also require fewer samples than machine learning approaches if the relationship between each feature and seizure occurrence is learned independently.

Many studies have aimed to predict, rather than forecast, seizures. Seizure prediction is binary, since the goal is to predict whether a seizure will or will not occur. This approach assumes that all seizures have a detectable pre-ictal state[29] and attempts to identify this state. However, when evaluating external risk factors, it is not pre-ictal states that are detected, but potential variations in seizure likelihood caused by these risk factors, thus implying a pro-ictal state[30]. In the pro-ictal state, seizures are more likely but not certain, and a probabilistic forecasting approach is suitable[31].

Materials and Methods

Dataset

Eight of fifteen patients from the NeuroVista seizure prediction clinical trial were analysed in this study[4]. The Human Research Ethics Committees of the participating institutes approved the NeuroVista trial and subsequent use of the data. All patients gave written informed consent before participation. Patient 3 was excluded due to high levels of signal dropout. Patients 2, 4, 5, 7, 12 and 14 were excluded because they had less than 50 seizures (lead and non-lead seizures during 0.5 – 2.0 years of recording). This cut-off was chosen to minimise unreliable seizure likelihood priors generated from patients with relatively few seizures. Patient demographics are shown in Table S2. Intracranial EEG (iEEG) recordings lasted for an average of 1.5 years per patient (range 1.0-2.1) with an average of 248 seizures per patient

(range 52-475). To allow for initial post-surgical instability of iEEG and seizure behaviours, only data from 100 days after implantation was used[32]. Events that were not clinically confirmed or electrographically similar to clinically confirmed events were excluded. Non-lead seizures, defined as seizures that followed another seizure by less than five hours, were also excluded to avoid potential confounding effects from an increased seizure chance during seizure clusters.

Table S1 shows the different features that were considered. To account for signal dropout, histograms for each feature were inspected and samples removed where values lay outside of expected ranges. For example, time since waking was sometimes greater than 24 hours due to overnight signal loss.

Sleep

The iEEG data were automatically labelled by classifiers trained according American Academy of Sleep Medicine 2012 methodology into rapid eye movement (REM) sleep, stage 1 sleep, stage 2 sleep, stage 3 sleep or awake. The iEEG data were scanned for artefacts using the algorithm of Nejedly et. al. 2019[33]. Each epoch that was not detected as artifactual was assigned into sleep categories based on methods adapted from Kremen et al. 2019[34] and Gerla et al. 2019[32]. This method was previously confirmed and validated on set of intracranial EEG data with concurrent polysomnography and gold standard sleep scoring according AASM2012 rules and yielded in average accuracy 94% with Cohen's kappa 0.87[34]. Sleep scoring was performed on a representative electrode or median of all electrodes driven by expert selection and reviewer judgement during manual review. For each patient, nine days at equidistant positions throughout the dataset were manually sleep-wake scored and used to train a patient-specific classifier. The trained classifier was then deployed in a daily manner, scoring 24-hour sections at a time. As a postprocessing step, days with insufficient or noisy data were classified as unknown if more than 50% of data in a day was missing. Subsequently, each 30-second epoch where more than 50% data was missing was set to unknown class.

From the classified data, nine objective sleep features were derived: current brain state, hours asleep, hours awake, hours in REM sleep, hours in stage 1, hours in stage 2 sleep, hours in stage 3 sleep, number of sleep stage transitions and time since waking. All (except time since waking and current brain state) were calculated from the preceding 24 hours. Time since waking was calculated from the previous sleep sample (this includes both night sleep and naps).

Where current brain state is classified as unknown, the previous known state (within an hour) was used.

Hours asleep and hours awake were not complementary because of iEEG signal dropout (median loss per patient: 0.125 to 7.125 hours per day, see supplementary Figure S4). Dropout will misrepresent seizure likelihoods if not accounted for. Samples were weighted according to the proportion of data lost. Thus, seizure likelihood for the j^{th} percentile bin of the i^{th} feature was calculated as

$$P(f_{i,j}) = \frac{\sum_{k=1}^{n}(1-d_k) * b_{i,j,k} * s_k}{\sum_{k=1}^{n}(1-d_k) * b_{i,j,k}}, \quad (1)$$

where $f_{i,j}$ represents the i^{th} feature value in the j^{th} percentile bin, n is the number of samples, d_k represents dropout in the k^{th} sample, $b_{i,j,k}$ represents the i^{th} feature value of the k^{th} sample that fell within the j^{th} percentile bin and s_k represents a seizure occurred in the k^{th} sample.

Weather

Hourly weather data was provided by the Australian Bureau of Meteorology. Temperature, humidity, wind speed, pressure and rainfall data were collected from the local weather station closest to each patient's home. Patients stayed near their home locations for the duration of the trial.

Maximum temperatures, minimum temperatures and pressure ranges were based on the 24 hours preceding the sample time. All other features used values associated with the hour preceding the sample time. Rain measures were simply whether there was any rainfall or not since most hours had no rain.

Temporal features

The times of seizures were binned according to hour of the day using 24-bin histograms. Similarly, weekday information was binned into the seven days of the week. To investigate monthly cycles, the lunar cycle was used as it is constant in length in contrast to calendar months. Some evidence suggests that the lunar cycle can affect sleep quality[36] and so may also affect seizure likelihood.

Forecasting seizure likelihood

Training and testing sets were allocated chronologically for pseudoprospective analysis. The first 50% of seizures were allocated to the training set and the remaining 50% of seizures to the testing set to ensure the widest possible range of weather in both sets.

The training and testing sets were segmented into consecutive 10-minute samples. These samples were aligned to the beginning of each set and not to seizure onset. Samples that contained the onset of a seizure were labelled "ictal" and all other samples were labelled "interictal". Each sample was considered a potential seizure occurrence period (SOP) and thus the SOP duration was 10 minutes. SOP was chosen to be 10 minutes as that is preferred by patients[37]. A 10-minute warning/intervention time was also used, matching median patient preference[38].

For each feature, histograms of seizure likelihood were generated (see Figure 1 and Supplementary Figures S1-3). Bin edges were defined by percentile values such that each bin contained 10% of the samples. Percentiles were used to ensure that each bin represented an equal number of samples, avoiding more extreme deviations of likelihood that arise when calculated using a bin with few samples.

For each sample in the test set, the forecast was given by the training set seizure likelihood associated with sample's weather, sleep or temporal feature value.

Combining forecasts

Forecasts from individual features were combined using the naïve Bayesian equation,

$$l = \frac{P(s) * \prod_i P(f_{i,j_i}|s)}{P(s) * \prod_i P(f_{i,j_i}|s) + P(s') * \prod_i P(f_{i,j_i}|s')}, \quad (2)$$

where l is the forecasted seizure likelihood given all included features, s is a seizure, s′ is the absence of a seizure, i is the current feature, j_i is the percentile bin of the i^{th} feature and $f_{i,j}$ represents that the i^{th} feature's value was within the j_i^{th} percentile bin.

Derivation of this equation is provided in Appendix S1. This equation assumes independence between all factors. This assumption is likely false but the naïve Bayesian method has been shown to work as an effective model even without independent variables[39]. Not relying on this assumption requires the calculation of a multivariate probability distribution, which was not feasible given the number of seizures and factors considered.

Measuring Performance

Many of the standard metrics of performance for seizure prediction, including sensitivity, false positive rate (FPR) and time in warning, do not apply for probabilistic forecasting because the binary distinction of true and false depend upon a set warning threshold. This issue was circumvented by producing Receiver-Operator Characteristic (ROC) curves, converting the forecasted probability to one or zero across varying thresholds. The Area Under the Curve (AUC) of the ROC curve is useful as a single-value measure of forecasting performance[5]. AUC assumes that discriminative thresholds are applied to a forecast to classify into either low or high risk. Other metrics that consider a continuous probability rather than a binary warning state[22] may be more appropriate to compare the clinical utility of seizure forecasts. The aim of this work was to evaluate the relative capability of sleep, weather and temporal features to discriminate pro-ictal (i.e. high risk) states from background (low risk) states linked to these features.

Samples in the hour following an ictal sample were considered post-ictal and so were not included when determining forecasting performance. Forecasters were considered to perform significantly better than chance if the AUC confidence interval did not contain 0.50 (the average value with surrogate time series forecasts). AUC confidence interval was calculated using the Hanley-McNeil method[40] ($\alpha = 0.05$ with Bonferroni correction). For details on the methodology for testing significance and surrogate time series see appendix 2.

Results

In the following sections, the term 'feature group' refers to a collection of features (e.g. sleep features). Figure 2 shows ROC curves for each patient in each of the feature groups (sleep, weather, temporal) and all feature groups combined. AUC values of the ROC curves are shown in Table 1. Six of eight patients had AUC significantly greater than chance level of 0.5 when combining all feature groups (average 0.684, range 0.503-0.770). For four of these six patients, combining all feature groups gave better results than forecasting with any one feature group.Temporal features performed better than chance for six patients, sleep features for five patients and weather features for two patients. Using only features that performed better than chance in the training set did not alter average test set performance (average AUC = 0.687, range 0.498-0.783). Per patient, the combined forecast outperforms the best performing individual feature group in four of the eight patients ($p=0.0495$, paired t-test).

Forecasting with sleep features

Figure 3 shows forecasting performance derived from objective measures of sleep (sleep features) in all patients. Forecasting with all sleep features combined performed better than chance for five out of eight patients, with an average of 0.630 (range 0.465-0.728) across all patients. Forecasting with the number of hours asleep, hours in REM sleep, hours in stage one sleep or the number of sleep stage transitions did not perform better than chance for any patient. Forecasting with the number of hours awake performed better than chance for two patients. Current state, hours in stage two and stage three sleep performed better than chance in one patient. Time since waking proved the most useful sleep feature, with better than chance forecasting in four patients.

Forecasting with weather features

Figure 4 shows forecasting performance for the weather features in all patients. Forecasting with combined weather features performed better than chance for two out of eight patients, with an average of 0.555 (range 0.336 – 0.698) across all patients. Forecasting with maximum temperature, minimum temperature, pressure and pressure range did not perform better than chance for any patient. Forecasting with humidity and wind speed performed better than chance for patient 9 only. Forecasting with rainfall performed better than chance for two patients. Forecasting with temperature at the time of the sample performed better than chance for three patients. For patient 6, temperature, maximum temperature and combining all weather features performed significantly worse than chance. This likely occurred due to the low seizure count resulting in a seizure likelihood distribution that was not representative of the patient's susceptibility to weather features.

Forecasting with temporal features

Figure 5 shows forecasting performance for the temporal features in all patients. Forecasting with all temporal features combined performed better than chance for six out of eight patients, with an average AUC of 0.694 (range 0.578-0.818) across all patients. Forecasting with the hour of the day performed better than chance for seven patients, averaging 0.699 (range 0.535-0.834) across all patients. Lunar phase performed better than chance in only patient six and day of the week did not perform better than chance for any patient.

For patients where both time of day and time since sleep perform better than chance, forecasting with these two features combined outperforms forecasts with either feature individually (mean improvement = 0.016, p=0.012, see supplementary table S5).

Discussion

This study showed the predictive power of environmental and physiological features that may be complementary to EEG, opening a new direction to search for effective factors in seizure forecasting. Sleep, weather and temporal feature groups all performed better than chance in different patients. No measure performed better than chance across all patients, further highlighting the patient-specific nature of seizure forecasting[6,7]. The combined forecast outperforms all individual feature groups in four of the eight patients.

Combining only features that performed above chance on training set data did not alter performance compared with using all features from all feature groups. This is likely because training set performance could not reliably predict test set performance. To validate forecasting performance of some features, dozens if not hundreds of seizures may need to be recorded over at least two years (to account for a full seasonal cycle in each set). The advantage of a Bayesian approach is that a weighted combination can be used so that even uncertain or weak prior information may contribute to improving an overall forecast. Therefore, patients may still derive early benefits from these features. We expect that performance would gradually improve as more data is collected and longer cycles are fully captured.

Patient 6 performed well in the training set but poorly in the test set, including an AUC significantly worse than chance when using weather. This reversal of the predictive power is likely due to seasonal changes in weather data. For example, susceptibility to hot days may have been missed in the training set as hot days only occurred in the test set. Patient 6 also has the least number of lead seizures (39) in the study. We conclude that patient 6 did not have enough seizures to reliably generate seizure likelihood distributions that were representative of the true distributions, leading to unreliable performance. This effect is likely present in all patients, causing training set performance to be an unreliable indicator of test set performance. Therefore, using the training set to select features should be avoided. A real-world implementation of these prior features could initially use all features from all feature groups and only alter feature sets when enough data is available to robustly compare performance.

Sleep

Time since waking was the best performing sleep feature (average AUC = 0.649, range 0.487 -0.769). Time since waking is related to time of day (average AUC = 0.699, range 0.535-0.834), but it was not clear which more directly influenced seizure likelihood. Where both are useful

individually, the combined forecast outperforms both features, indicating that the two features have distinct utility to some degree.

For people with an irregular sleep schedule (see supplementary figure S6), time since waking may be more informative than time of day. Time since waking has a value of zero while asleep and so infers current brain state (sleep vs awake) which may also contribute to its performance. Further research is needed to explore the relationship between sleep, time and seizures.

Current brain state only performed better than chance in two patients. Current brain state does not consider multiple samples, like the other sleep features and therefore is more heavily impacted by unknown sleep state which may have led to this result.

For all patients, most sleep features other than time since waking did not perform better than chance. This was unexpected as sleep deprivation is associated with increased seizure frequency[25,26] although this has recently been contradicted[27] and the relationship is less clear when sleep deprivation is milder[41,42].

Though smaller day-to-day variations in sleep may modulate seizure likelihood, our results indicate it may not be useful as a seizure forecasting feature. The duration of stage two and stage three sleep was predictive of seizures for one patient, which aligns with previous findings that epileptic activity is mainly facilitated by non-REM sleep[43].

Signal loss and the resulting unknown vigilance state limit the conclusions we can make from this data (see Appendix S4). This has been partially accounted for by weighing the importance of samples by signal loss (see eqn. 1), although the effects of lost data cannot be fully counteracted. Patients 1 and 9 had large parts of the night classified as unknown (see supplementary Figure S5) which may explain their poorer performance using sleep features.

In this study, sleep features were measured from iEEG; however, some of these sleep features may eventually be able to be measured non-invasively. Further technological improvements are necessary before non-invasive devices, such as wearables, can be used to accurately measure sleep stages compared to gold standard polysomnography. Interestingly, the most useful sleep features (i.e. time since waking) also require the least precision and so may be utilized without an invasive device

Weather

Weather features showed the worst forecasting performance, with only two patients showing better than chance performance when combining features and only seven of 64 patient-feature

pairings showing significant results. Considering that patients were likely indoors for much of the study, it is promising that weather was even slightly predictive of seizures. Patient-localized sensors may improve performance of features such as temperature, pressure, and humidity. For most people with epilepsy, weather may simply not be a factor that contributes to seizure likelihood as only one of eight patients showed a positive result. For others the effect is small and so weather will likely play a minor role in forecasting, though still potentially useful.

Current temperature, humidity, windspeed and rainfall all performed better than chance in at least one patient, but pressure, pressure range, maximum temperature and minimum temperature did not. Changes in atmospheric pressure can affect seizure rate[44] although perhaps higher temporal resolution is required for pressure change to have utility.

The results indicate that current weather information was more useful than recent weather information. High humidity was predictive of increased seizure likelihood in a previous study of an in-patient population[24], and the current results (see Figure S2) support that this effect is also true for individual patients.

Temporal Features

Time of day was the most reliable of all individual features with all but patient 6 performing better than chance. This is not surprising as time of day paces the endogenous circadian cycle which is known to influence seizure timing[45]. This may be because time of day is also a useful proxy for both weather conditions and wakefulness. Given the strength of circadian features[22,23], it is reasonable to conclude that time of day should simply have been used alone without using weather or sleep data at all. This choice may be correct for some patients since weather and sleep may not significantly affect their seizure likelihood, although many patients are likely to be interested in understanding their personal seizure risk factors in addition to receiving a "black-box" forecast. Furthermore, the current results show that selecting single or a few features is generally less reliable than using combined forecasts.

Only patient 6 performed better than chance with lunar cycles and no patient showed significant performance with weekly cycles. Patient-specific multidien cycles from iEEG have been shown[23], including in these patients[46] and have produced robust prior probabilities[47]. However, patient-specific cycles may not always be known and so temporal cycles were restricted to fixed calendar-based cycles. Misalignment between patient-specific and calendar cycles likely led to the poor performance of multidien temporal features shown. The mismatch in timescales between forecaster and data may also reduce performance.

Using time since last seizure is perhaps an obvious candidate for a temporal feature, especially given seizures occur in bursts. However, the feature was omitted as only lead seizures were considered which could undervalue the feature's utility.

Forecasting

The Bayesian method of combining forecasts falsely assumes independence between each of the features being combined. Ideally, a multivariate likelihood distribution would be generated to account for the complex relationships between features. This would require an impractical number of seizures with a further increase in numbers required with each additional feature. By assuming independence, the number of seizures needed becomes feasible while still producing useful forecasts. This study cannot determine if the tested features are independent from each other or features not tested such as patient-specific multidien cycles.

Seizure number is still a limiting factor on whether forecasting in this framework is achievable. As seen with patient 6, when too few seizures were used to train the forecaster, performance was unreliable. This was largely due to overfitting of data in the training set. Overfitting may have been reduced by reducing the number of histogram bins. However, fewer bins means that is that it becomes harder to observe changes in seizure likelihood due to extreme conditions.

Ten feature bins were used as a trade-off between reliability and sensitivity to extreme events. The effects of extreme events were still likely diluted due to their rarity. Some people may have been highly susceptible to seizures in extreme scenarios (such as no sleep in 24 hrs or a 45-degree Celsius day) but these scenarios were not observed frequently enough during the recording period. For these rare but potentially high-risk scenarios, an alternative method for determining seizure probability, such as conditioning on the likelihood of rare events or even using population-wide histograms, may be more appropriate. Further research should investigate metrics for quantifying seizure probability under extreme scenarios.

Some people have multiple and distinct seizure populations distinguished by seizure length[48,49]. It is possible that these populations react differently to external factors and therefore accounting for different seizure populations could improve forecasting performance. In addition, if seizure populations can be distinguished, external factors could enable not only a forecast of when a seizure occurs, but also the anticipated severity of the seizure.

This paper developed seizure forecasts from three groups of features that could be measured non-invasively. While the set of possible features that was explored was limited by the available data, there are likely to be more features that will prove beneficial for seizure forecasting, such as heart rate[50,51], cortisol levels, alcohol consumption, or medication compliance[2]. As more features are explored, more patients may see significant performance or performance improvements.

Conclusion

This study builds on our previous work to show the potential power of combining new environmental and behavioural features. Although EEG signal features will likely continue to outperform auxiliary signals in seizure forecasting, the presented results provide complementary features that can be used in addition to invasively measured features or when such recordings are not available. Ultimately, it is our hope that a better understanding of patient-specific risk factors is a step towards making seizure forecasting a clinical reality for people with epilepsy.

Acknowledgements

The authors acknowledge the support of the National Health and Medical Research Council, Project Grant ID 1065638. Daniel Payne acknowledges the support of a Melbourne Research Scholarship, The University of Melbourne and support provided by Mentone Grammar. Vaclav Kremen and Vaclav Gerla were partially supported by Institutional Resources of Czech Technical University in Prague. Vaclav Kremen and Gregory A. Worrell were supported by National Institutes of Health under Grant R01 NS09288203 and Grant UH2/UH3NS95495.

Data Availability Statement

The NeuroVista Kaggle competition and seizure data are accessible online via https://www.epilepsyecosystem.org.

Disclosure of Conflicts of Interest

None of the authors has any conflict of interest to disclose

Ethical Publication Statement

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

References

1. Epilepsy Foundation. Ei2 community survey [Internet]. 2016 [cited 2018 Dec 14]. Available from: https://www.epilepsy.com/sites/core/files/atoms/files/community-survey-report-2016 V2.pdf

2. Dumanis SB, French JA, Bernard C, et al. Seizure Forecasting from Idea to Reality . Outcomes of the My Seizure Gauge Epilepsy Innovation Institute Workshop. eNeuro. 2017;2(December):1–5.

3. Fisher RS, Vickrey BG, Gibson P, et al. The impact of epilepsy from the patient ' s perspective I . Descriptions and subjective perceptions. Epilepsy Research. 2000;41:39–51.

4. Cook MJ, Brien TJO, Berkovic SF, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy : a first-in-man study. The Lancet Neurology. 2013;12(6):563–71.

5. Kuhlmann L, Karoly P, Freestone DR, et al. Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial EEG. Brain. 2018;(August):1–13.

6. Kuhlmann L, Lehnertz K. Seizure prediction — ready for a new era. Nature Reviews Neurology. 2018;14(october).

7. Freestone DR, Karoly PJ, Cook MJ. A forward-looking review of seizure prediction. Current Opinion in Neurology. 2017;30(2):1.

8. Bedeeuzzaman M, Fathima T, Khan YU, et al. Seizure prediction using statistical dispersion measures of intracranial EEG. Biomedical Signal Processing and Control. 2013;10:338–41.

9. Costa RP, Oliveira P, Rodrigues G, et al. Epileptic Seizure Classification Using Neural Networks with 14 Features. In: Lovrek I, Howlett RJ, Jain LC, editors. Knowledge-Based Intelligent Information and Engineering Systems: 12th International Conference, KES 2008, Zagreb, Croatia, September 3-5, 2008, Proceedings, Part II. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 281–8.

10. Winterhalder M, Schelter B, Maiwald T, et al. Spatio-temporal patient-individual assessment of synchronization changes for epileptic seizure prediction. Clinical

Neurophysiology. 2006;117(11):2399–413.

11. Mormann F, Andrzejak RG, Elger CE, et al. Seizure prediction : the long and winding road. Brain. 2007;130:314–33.

12. Park Y, Luo L, Parhi KK, et al. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. Epilepsia. 2011;52(10):1761–70.

13. Le Van Quyen M, Soss J, Navarro V, et al. Preictal state identification by synchronization changes in long-term intracranial EEG recordings. Clinical Neurophysiology. 2005;116(3):559–68.

14. Ghaderyan P, Abbasi A, Sedaaghi MH. An efficient seizure prediction method using KNN-based undersampling and linear frequency measures. Journal of Neuroscience Methods. 2014;232:134–42.

15. Schelter B, Feldwisch-Drentrup H, Ihle M, et al. Seizure prediction in epilepsy: From circadian concepts via probabilistic forecasting to statistical evaluation. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2011;1624–7.

16. Parvez MZ, Paul M. Seizure Prediction using Undulated Global and Local Features. IEEE Transactions on Biomedical Engineering. 2016;9294(c):1–1.

17. Kuhlmann L, Freestone D, Lai A, et al. Patient-specific bivariate-synchrony-based seizure prediction for short prediction horizons. Epilepsy Research. 2010;91(2–3):214–31.

18. Mirowski P, Madhavan D, Le Cun Y, et al. Classification of patterns of EEG synchronization for seizure prediction. Clinical Neurophysiology. 2009;120(11):1927–40.

19. Kiral-Kornek I, Roy S, Nurse E, et al. Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System. EBioMedicine. 2017;

20. Spalt J, Langbauer G, Boltzmann L, et al. Subjective perception of seizure a questionnaire study precipitants : results of. Seizure. 1998;7.

21. Bercel NA. The periodic features o f some seizure states. Annals of the New York Academy of Sciences. 1964;

22. Karoly PJ, Ung H, Grayden DB, et al. The circadian profile of epilepsy improves seizure forecasting. Brain. 2017;140(8):2169–82.

23. Baud MO, Kleen JK, Mirro EA, et al. Multi-day rhythms modulate seizure risk in epilepsy. Nature Communications. 2018;9:1–10.

24. Rakers F, Walther M, Schiffner R, et al. Weather as a risk factor for epileptic seizures : A case-crossover study. Epilepsia. 2017;58(7):1287–95.

25. Ellingson RJ, Wilken K, Bennett DR. Efficacy of sleep deprivation as an activation procedure in epilepsy patients. Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society. 1984 Jan;1(1):83—101.

26. Foldvary-Schaefer N, Grigg-Damberger M. Sleep and epilepsy: what we know, don't know, and need to know. Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society. 2006 Feb;23(1):4—20.

27. Rossi KC, Joe J, Makhija M, et al. Insufficient Sleep, Electroencephalogram Activation, and Seizure Risk : Re-Evaluating the Evidence. Annals of Neurology. 2020;87(6):798–806.

28. Russel SJ, Norvig P. Bayes rule and its use. In: Hirsch M, editor. Artificial intelligence A modern approach. 3rd ed. New Jersey: Pearson; 2010. p. 495–9.

29. Freestone DR, Karoly PJ, Peterson ADH, et al. Seizure Prediction: Science Fiction or Soon to Become Reality? Curr Neurol Neurosci Rep. 2015;15:73.

30. Baud MO, Proix T, Rao VR, et al. Chance and risk in epilepsy. Current Opinion in Neurology. 2020;33(2):163–72.

31. Litt B, Lehnertz K. Seizure prediction and the preseizure period. Current Opinion in Neurology. 2002;(May).

32. Ung H. Intracranial EEG Fluctuates Over Months After Implanting Electrodes in Human Brain. Journal of Neural Engineering. 2017;14(5):056011.

33. Nejedly P, Kremen V, Sladky V, et al. Exploiting Graphoelements and Convolutional Neural Networks with Long Short Term Memory for Classification of the Human Electroencephalogram. 2019;(July):2–10.

34. Kremen V, Brinkmann BH, Van Gompel JJ, et al. Automated unsupervised behavioral

state classification using intracranial electrophysiology. Journal of Neural Engineering. 2019;16(2):1–8.

35. Gerla V, Kremen V, Macas M, et al. Iterative expert-in-the-loop classification of sleep PSG recordings using a hierarchical clustering. Journal of Neuroscience Methods. 2019;317:61–70.

36. Cajochen C, Frey S, Knoblauch V. Report Evidence that the Lunar Cycle Influences Human Sleep. 2013;1485–8.

37. Schulze-Bonhage A, Sales F, Wagner K, et al. Views of patients with epilepsy on seizure prediction devices. Epilepsy and Behavior. 2010;18(4):388–96.

38. Arthurs S, Zaveri HP, Frei MG, et al. Patient and caregiver perspectives on seizure prediction. Epilepsy and Behavior. 2010;19(3):474–7.

39. Domingos P, Pazzani M. Beyond Independence : Conditions for the Optimality of the Simple Bayesian Classi er. In: 13th Intl Conf Machine Learning. 1996.

40. Hanley JA, Mcneil BJ. A Method of Comparing the Areas under Reveiver Operating Characteristic Curves Derived from the Same Cases '. Radiology. 1983;148(3):839–43.

41. Cobabe MM, Sessler DI, Nowacki AS, et al. Epilepsy & Behavior Impact of sleep duration on seizure frequency in adults with epilepsy : A sleep diary study. Epilepsy & Behavior. 2015;43:143–8.

42. Samsonsen C, Sand T, Bråthen G, et al. The impact of sleep loss on the facilitation of seizures : A prospective case-crossover study. Epilepsy Research. 2016;127:260–6.

43. Frauscher B, von Ellenrieder N, Ferrari-Marinho T, et al. Facilitation of epileptic activity during sleep is mediated by high amplitude slow waves. Brain. 2015;138:1629–41.

44. Doherty MJ, Youn C, Gwinn RP, et al. Atmospheric Pressure and Seizure Frequency in the Epilepsy Unit : Preliminary Observations. Epilepsia. 2007;48(9):1764–7.

45. Quigg M, Straume M. Dual epileptic foci in a single patient express distinct temporal patterns dependent on limbic versus nonlimbic brain location. Annals of Neurology. 2001;48(1):117–20.

46. Karoly PJ, Goldenholz DM, Freestone DR, et al. Circadian and circaseptan rhythms in human epilepsy : a retrospective cohort study. The Lancet Neurology. 2018;17(11):977–

85.

47. Maturana MI, Meisel C, Dell K, et al. Critical slowing as a biomarker for seizure susceptibility. Nature Communications. 2020;11(2172).

48. Cook MJ, Karoly PJ, Freestone DR, et al. Human focal seizures are characterized by populations of fixed duration and interval. Epilepsia. 2016;57(3):359–68.

49. Payne DE, Karoly PJ, Freestone DR, et al. Postictal suppression and seizure durations: A patient-specific, long-term iEEG analysis. Epilepsia. 2018;59(5):1027–36.

50. Kerem DH, Geva AB. Forecasting epilepsy from the heart rate signal. Medical and Biological Enginineering and Computing. 2005;43(1981).

51. Fujiwara K, Miyajima M, Yamakawa T, et al. Epileptic Seizure Prediction Based on Multivariate Statistical Process Control of Heart Rate Variability Features. IEEE Transactions on Biomedical Engineering. 2015;9294(c):1–1.

Figure Legends

Figure 1: Example of a seizure likelihood histogram and heatmap. A) Seizure likelihood across a range of humidity levels for patient 8. Each histogram bin represents 10% of all samples and so they are not of consistent width. Both the height and color represent seizure likelihood. B) Histograms for each patient are converted into a row of the heatmap, keeping the color representation and width shown in the histograms.

Figure 2: Receiver operator characteristic curves for all patients using A) combined temporal features, B) combined weather features, C) combined sleep features and D) all feature groups combined. E) AUC values for patient 15 exemplifying how performance can improve when features are combined.

Figure 3: AUC scores when forecasting using sleep features in the A) training set and B) testing set. Each section represents a different sleep feature. The final section shows scores when the forecasts from all sleep features were combined. Average lines indicate the average for that feature across patients. Chance performance is 0.5 as indicated by the thin line. Error bars represent the upper limits of the confidence interval of each AUC score.

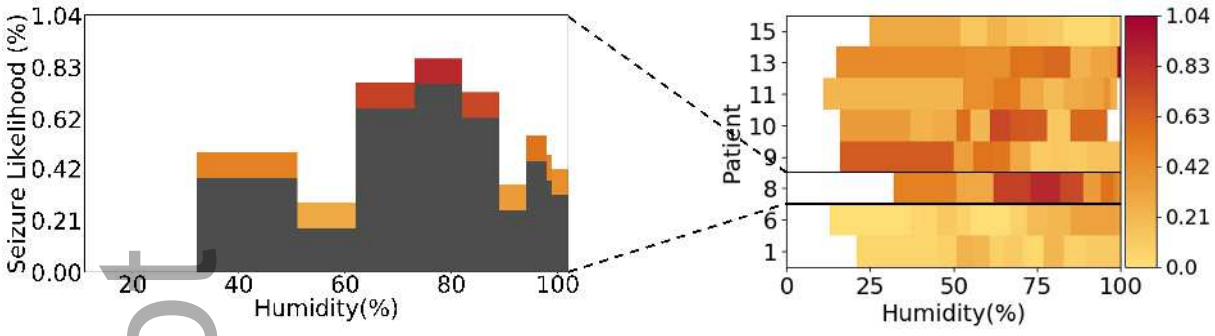
Figure 4: AUC scores when forecasting using weather features in the A) training set and B) testing set. Each section represents a different weather feature. The final section shows scores when the forecasts from all weather features were combined. Average lines indicate the average for that feature across patients. Chance performance is 0.5 as indicated by the thin line. Error bars represent the upper limits of the confidence interval of each AUC score.

Figure 5: Comparison of AUC scores when forecasting using temporal features in the A) training set and B) testing set. Each section represents a different temporal feature. The final section shows scores when the forecasts from all temporal features were combined. Average lines indicate the average for that feature across patients. Chance performance is 0.5 as indicated by the thin line. Error bars represent the upper limits of the confidence interval of each AUC score.

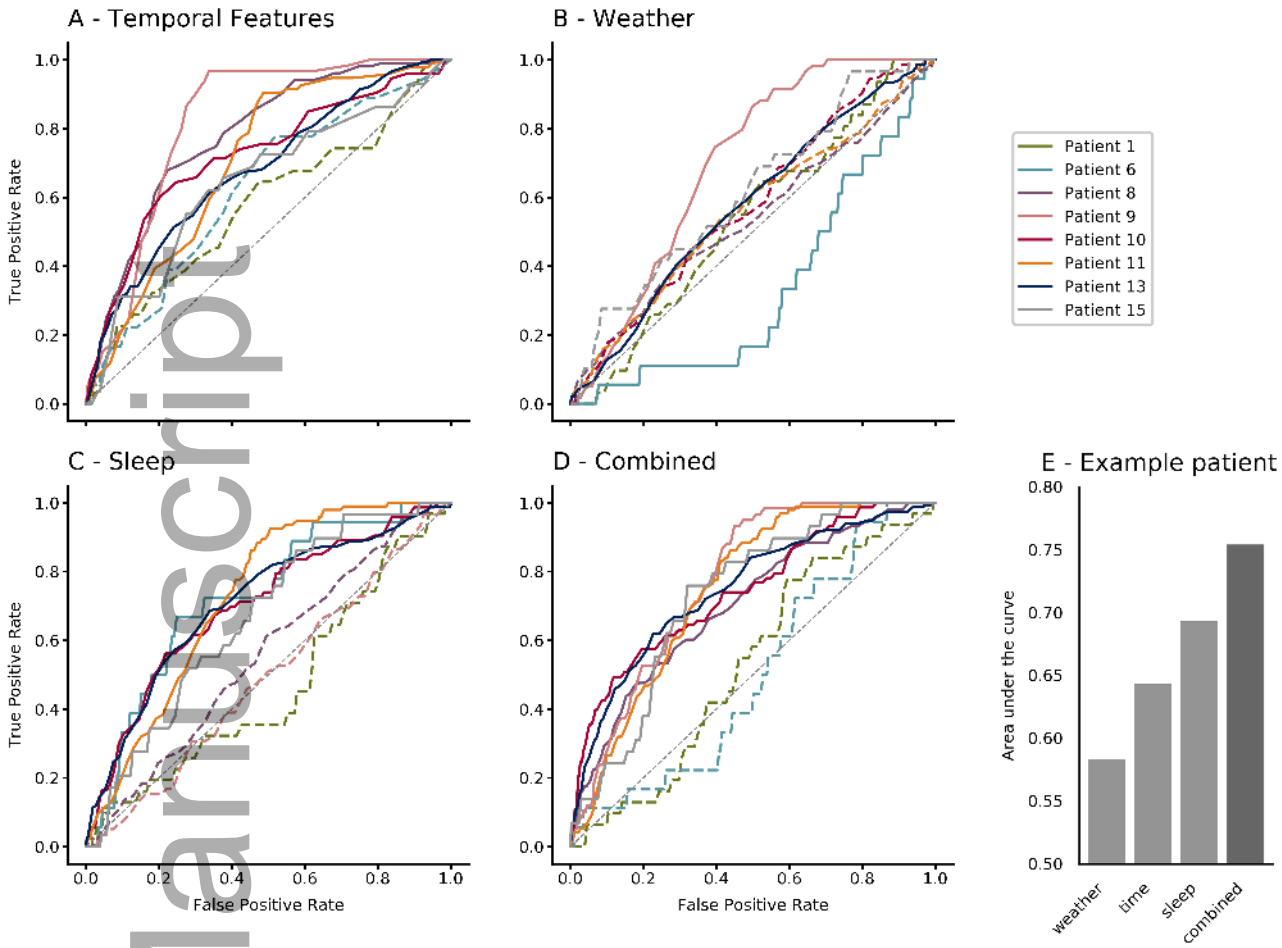
Table 1: Test set area under the curve scores across patients and feature groups.

Patient	Seizures	P(Sz)	Sleep	Weather	Temporal	Combined
1	74	0.000801	0.465(p=0.487)	0.538 (p=0.47)	**0.578 (p=0.15)**	0.531 (p=0.55)
6	39	0.000794	**0.722* (p=0.001)**	0.336* (p=0.003)	0.625 (p=0.077)	0.503 (p=0.96)
8	224	0.00339	0.555 (p=0.062)	0.538 (p=0.20)	**0.780* (p<0.001)**	0.702* (p<0.001)
9	146	0.00343	0.492 (p=0.82)	0.698* (p<0.001)	**0.818* (p<0.001)**	0.770* (p<0.001)
10	168	0.00427	0.701* (p<0.001)	0.578 (p=0.027)	0.721* (p<0.001)	**0.743* (p<0.001)**
11	212	0.00237	0.728* (p<0.001)	0.558 (p=0.059)	0.704* (p<0.001)	**0.747* (p<0.001)**
13	305	0.00351	0.709* (p<0.001)	0.576* (p=0.002)	0.680* (p<0.001)	**0.747* (p<0.001)**
15	58	0.00110	0.664* (p=0.003)	0.616 (p=0.037)	0.644* (p<0.001)	**0.731* (p<0.001)**

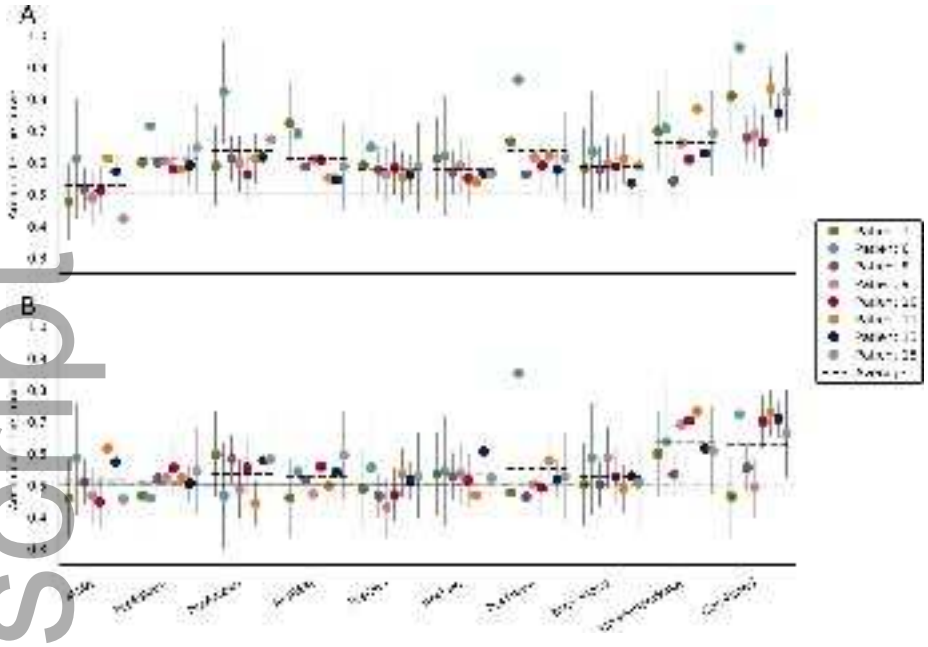
* - chance performance (0.5) is not within the confidence interval for the area under the curve (α=.0125 after Bonferroni correction). Bold text indicates the best performance for each patient. P(Sz) – Probability of seizure in a 10-minute sample, calculated from the combined train and test data.



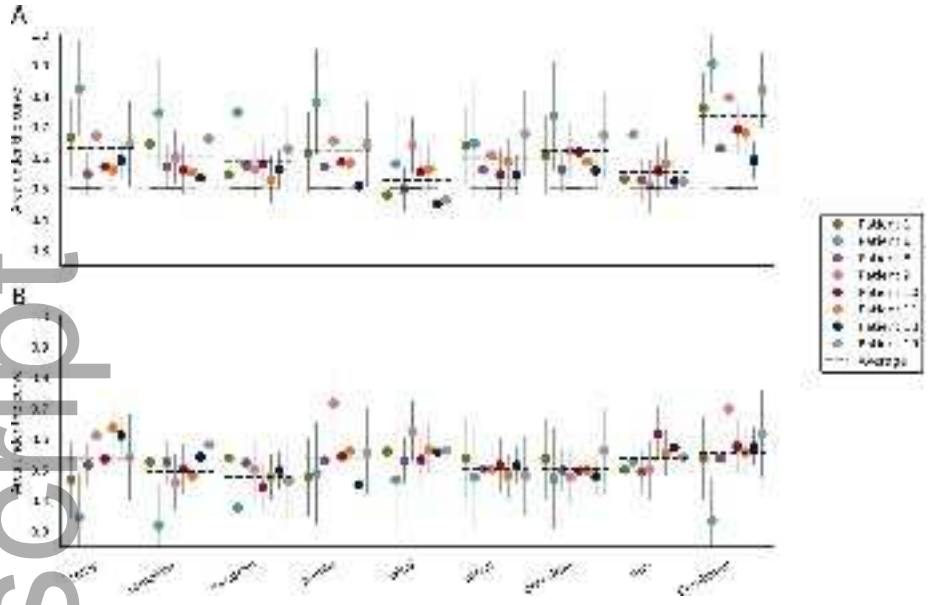
epi_16785_f1.tif



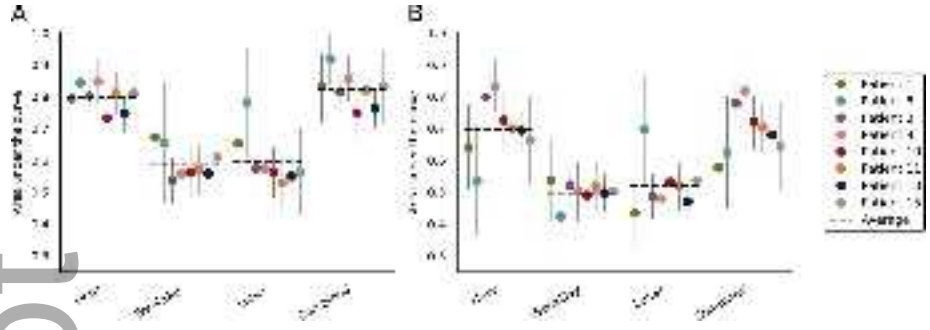
epi_16785_f2.tif



epi_16785_f3.tif



epi_16785_f4.tif



epi_16785_f5.tif