Seizure prediction with spectral power of EEG using cost-sensitive support vector machines

*Yun Park, *Lan Luo, *Keshab K. Parhi, and †Theoden Netoff

*Departments of *Electrical Engineering and †Biomedical Engineering, University of Minnesota, Minneapolis, Minnesota, U.S.A.*

SUMMARY

Purpose: We propose a patient-specific algorithm for seizure prediction using multiple features of spectral power from electroencephalogram (EEG) and support vector machine (SVM) classification.

Methods: The proposed patient-specific algorithm consists of preprocessing, feature extraction, SVM classification, and postprocessing. Preprocessing removes artifacts of intracranial EEG recordings and they are further preprocessed in bipolar and/or time-differential methods. Features of spectral power of raw, or bipolar and/or time-differential intracranial EEG (iEEG) recordings in nine bands are extracted from a sliding 20-s–long and half-overlapped window. Nine bands are selected based on standard EEG frequency bands, but the wide gamma bands are split into four. Cost-sensitive SVMs are used for classification of preictal and interictal samples, and double cross-validation is used to achieve in-sample optimization and out-of-sample testing. We postprocess SVM classification outputs using the Kalman Filter and it removes sporadic and isolated false alarms. The algorithm has been tested on iEEG of 18 patients of 20 available in the Freiburg EEG database who had three or more seizure events. To investigate the discriminability of the features between preictal and interictal, we use the Kernel Fisher Discriminant analysis.

Key findings: The proposed patient-specific algorithm for seizure prediction has achieved high sensitivity of 97.5% with total 80 seizure events and a low false alarm rate of 0.27 per hour and total false prediction times of 13.0% over a total of 433.2 interictal hours by bipolar preprocessing (92.5% sensitivity, a false positive rate of 0.20 per hour, and false prediction times of 9.5% by time-differential preprocessing). This high prediction rate demonstrates that seizures can be predicted by the patient-specific approach using linear features of spectral power and nonlinear classifiers. Bipolar and/or time-differential preprocessing significantly improves sensitivity and specificity. Spectral powers in high gamma bands are the most discriminating features between preictal and interictal.

Significance: High sensitivity and specificity are achieved by nonlinear classification of linear features of spectral power. Power changes in certain frequency bands already demonstrated their possibilities for seizure prediction indicators, but we have demonstrated that combining those spectral power features and classifying them in a multivariate approach led to much higher prediction rates. Employing only linear features is advantageous, especially when it comes to an implantable device, because they can be computed rapidly with low power consumption.

KEY WORDS: Epilepsy, Seizure, Prediction, Detection, Electroencephalogram.

Recently, there has been great progress in seizure suppression methods. Deep brain stimulation therapy has been demonstrated to abate seizures in clinical trials (Fisher et al., 2010). More experimental approaches have been used to suppress seizures by optically uncaging inhibitory neurotransmitters (Yang et al., 2009) or by focal cooling of the cortex (Rothman et al., 2005). The efficacy of these methods may be improved by a closed-loop therapy, where a seizure prediction device monitors and triggers the seizure abatement. However, no seizure prediction algorithm has yet been developed that has sufficient sensitivity and specificity and can be implemented in an implantable device (Mormann et al., 2007).

Seizure prediction using electroencephalogram (EEG) with high sensitivity and specificity has been elusive, despite numerous claims that a proposed algorithm or measure has provided significant predictive power. For example, nonlinear measures taken from chaos theory and applied to intracranial EEG (iEEG) demonstrated promising predictive power (Martinerie et al., 1998). However, when compared to linear features, the nonlinear features were not significantly better (Mormann et al., 2005) and their

computational intensiveness made them prohibitive to be calculated in real-time. Furthermore, promising nonlinear features were not predictive at all when tested on long time series (Harrison et al., 2005).

Seizure prediction based on EEG/iEEG is complicated by two factors. The first is that preictal and interictal EEG/iEEG patterns across patients vary substantially. There may be no single generic algorithm that can be applied to all patients and can achieve high sensitivity and specificity (Osorio et al., 1998; Shoeb et al., 2009). The second is that EEG/iEEG is highly complex and varies over time, and no single measure of EEG/iEEG has yet been predictive on its own (Mormann et al., 2005, 2007; Feldwisch-Drentrup et al., 2010). Therefore, we hypothesize that a *patient-specific* classification method based on *multiple* features extracted from EEG/iEEG will achieve high sensitivity and specificity.

Our patient-specific approach to seizure prediction is based on binary classification of iEEG using a machine learning algorithm. A machine learning algorithm classifies samples of iEEG as either preictal (immediately prior to a seizure) or interictal (between seizures) based on their multivariate features (see Fig. 1). When an epoch of iEEG is classified as preictal, the device can trigger an alarm or a seizure prevention device. In our study, the algorithm is trained by providing it with iEEG labeled by a clinician as preictal and interictal, and tested in iEEG data that have not been touched in training. We hypothesize that machine learning approaches will work well for seizure prediction in patients who have stereotypical recurrent seizures.

Of the available classifiers we have chosen support vector machines (SVMs) (Vapnik, 2000; Cherkassky & Mulier, 2007). The SVM is a margin-based classifier that maps input data onto a high-dimensional space and classifies them with a linear approximation (see Fig. S1). The SVM is considered the most powerful and favorable classifier in the statistical learning community (Alpaydin, 2004; Bishop, 2006; Cherkassky & Mulier, 2007). We use a cost-sensitive SVM (CSVM) that has an option of different misclassification costs for different groups, because preictal data are scarcer than the interictal data and more important to identify.

In many previous studies the optimization of the classifier is done with the "test set," which may result in an unintended influence of the test data on the classifier thus higher sensitivity and specificity than those that may actually be experienced in real-world conditions (Mormann et al., 2007). To provide more actual estimates of sensitivity and specificity, the test data must not influence the training of the algorithm. Therefore, we employ a *double cross-validation* method, where data are divided into a *training* set and a *test* set, and the training set is further subdivided into a *learning* set and a *validation* set (Friedman, 1994; Cherkassky & Mulier, 2007). The SVM-classification model is trained on learning and validation datasets and tested on a dataset that is not touched in training. Because the test dataset is left completely out of training, the results with this experimental design can more accurately reflect the expected prediction rate in real-world conditions.

Our ultimate goal is to develop algorithms and architectures for an implantable device that can reliably provide seizure prediction with sufficient time to trigger an antiepileptic therapy. The specific task in this study is to investigate the feasibility of a patient-specific classification approach with the CSVM to distinguish between preictal and interictal iEEG using only linear features. Linear features are advantageous for an implantable device, where seizure prediction must be done in real time with constraints of power consumption. The algorithm has been tested on data from the Freiburg EEG database, which has been made available for comparing the results of different algorithms on the same datasets. Sensitivity and false positive rates are reported for 18 patients selected from the database. Patients who had fewer than three seizures (the minimum necessary for double cross-validation) or had no available interictal recordings were excluded from this study.
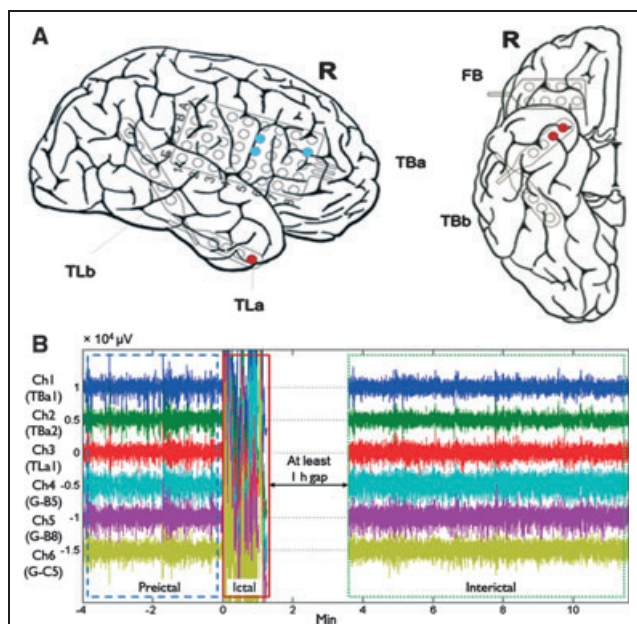

Figure 1.
Electrode positions and iEEG recorded on them with a seizure event. (**A**) Three electrodes placed near focal (red, close to the seizure origination) and the other three near afocal (blue, away from focal areas). The figures are provided along with Patient 17's iEEG in the Freiburg database. (**B**) iEEG with a seizure event. Ictal (seizure, indicated by solid red box) is immediately preceded by a window we define as "preictal" (indicated by dashed blue box). Windows at least an hour prior to or after a seizure are defined as "interictal" (indicated by dotted green box) and are assumed to represent ordinary iEEG activity.
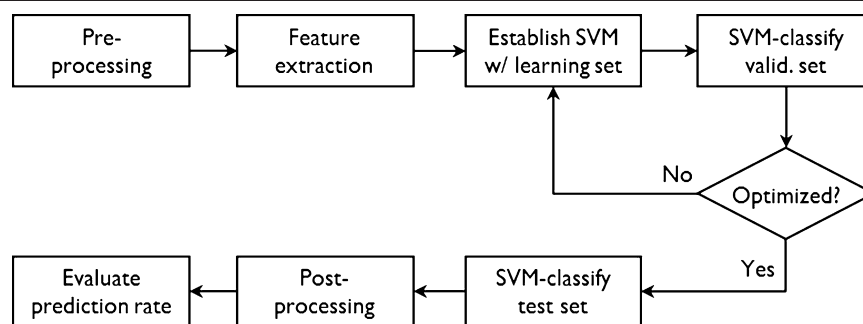*Epilepsia* © ILAE

Figure 2.
Outline of the proposed seizure prediction algorithm.
Epilepsia © ILAE

METHODS

The proposed algorithm consists of preprocessing, feature extraction, SVM classification using double cross-validation, and postprocessing (see Fig. 2). Each component of the proposed algorithm is discussed in detail in its corresponding subsection below.

Patient database

We have trained and tested our algorithm on the Freiburg EEG database (https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database), which is available to any lab by request. This database contains electrocorticogram (ECoG) or iEEG from 21 patients with medically intractable focal epilepsy. We have chosen 18 of the available datasets of 20 patients, who have three or more seizures (the minimum number for double cross-validation). Each 20-s–long window of iEEG has been categorized as ictal (containing a seizure), interictal (at least 1 h preceding or postceding a seizure), preictal (in 30 min preceding a seizure onset), or artifact. Half an hour of iEEG recordings preceding preictal and an hour of those postceding seizure offset are excluded in training. The Freiburg database contains six of iEEG recordings from grid, strip, or depth-electrodes, three near the seizure focus (focal) and the other three distal to the focus (afocal). Seizure onset times and artifacts were identified by certified epileptologists. The data were collected at 256 Hz (Patient 12's interictal at 512 Hz) sampling rate with 16 bit analog-to-digital converters.

Preprocessing: removing artifacts of iEEG recordings

iEEG data are subject to artifacts, such as line noise, electrical noise, and movement artifacts. Many of these artifacts may distort original iEEG and affect the further process of training and testing. Therefore, iEEG recordings with artifacts are removed from further analysis. Artifacts in the Freiburg iEEG recordings were marked by epileptologists, and the 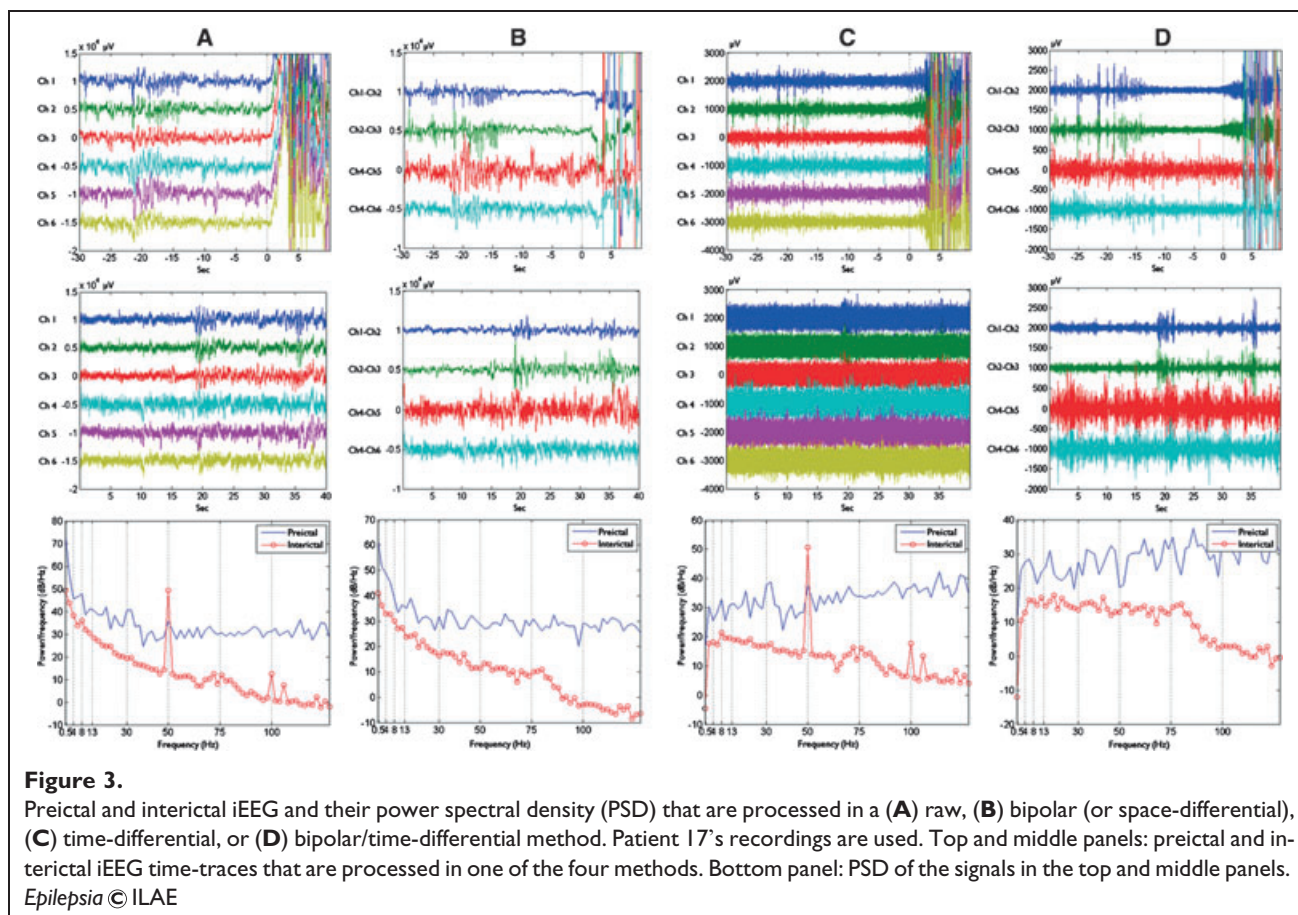information about the artifacts is provided along with the datasets. We have removed windows containing those artifacts, and the proportion of the removed artifact windows to the overall recordings is negligible (approximately 10 min in aggregate). Power line hums at 50 and 100 Hz have been removed by excluding spectral power in the bands of 47–53 and 97–103 Hz when the features are extracted.

In addition, bipolar and/or time-differential methods have been used to remove or reduce the effect of other types of artifacts in iEEG (see Fig. 3). The bipolar (or space-differential) measurement provides common-mode rejection to reduce line noise and movement artifacts that are common to all the electrodes. The bipolar recording method is commonly used in the analysis of EEG and provides better spatial resolution than that of ordinary reference recordings (Nunez & Srinivasan, 2006). In our study, bipolar electrode recordings are made preferentially between channels within focal or afocal and only between the two groups if they are physical neighbors. This results in 4–6 bipolar recordings for each patient.

In general, raw iEEG shows much more power in low frequency bands than in high frequency bands, making it difficult to compare power across the bands. We normalize the power in each band by measuring its contribution to the total power, but the normalized power is dominated by small changes in power in low frequency bands. Therefore, the proportion of high frequency power in the total power is influenced by low frequency power. The time-differential method (an approximate derivative in signals, $d[n] = x[n] - x[n-1]$) provides a way to reduce that undesired effect by flattening the spectrum, making power in the high frequency bands similar to that in low frequency bands. The time-differential processing is also known as the Hjorth mobility parameter (Hjorth, 1970).

Feature extraction

We have calculated spectral power in nine bands in a 20-s–long window of iEEG and used it as a feature set in this study. We adopted the moving window analysis (Mormann

Figure 3.
Preictal and interictal iEEG and their power spectral density (PSD) that are processed in a (**A**) raw, (**B**) bipolar (or space-differential), (**C**) time-differential, or (**D**) bipolar/time-differential method. Patient 17's recordings are used. Top and middle panels: preictal and interictal iEEG time-traces that are processed in one of the four methods. Bottom panel: PSD of the signals in the top and middle panels.
Epilepsia © ILAE

et al., 2007) with a half overlap and it provides a prediction of a seizure every 10 s based on the analysis of 5,120 time points. Spectral bands are selected based on standard iEEG frequency bands, but the wide gamma band is split into four bands (Netoff et al., 2009; Park et al., 2010): delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), four gamma bands (30–47, 53–75, 75–97, and 103–128 Hz, excluding power line hums), and their total. Power in each of the above bands is divided by the total power and the last feature included is the total power (Litt & Echauz, 2002). iEEG data from each of the six electrodes are broken into 20-s windows that half-overlap the previous one, and nine spectral features are extracted from each window. A total of 54 features are extracted from every 10 s of raw or time-differential iEEG (36–48 features from bipolar or bipolar/time-differential). Because we have four ways to preprocess iEEG, we will compare seizure prediction based on each of the four preprocessing methods: raw, bipolar, time-differential, and bipolar/time-differential.

SVM classification using double cross-validation

Our classification task is to distinguish between two groups of iEEG data features: preictal and interictal. In our algorithm, the binary classification is performed in two steps: SVM classification with the 20-s window and post-

processing using the Kalman filter (postprocessing will be explained in the next subsection in detail). The goal of SVM classification is establishing and testing a mapping $x \rightarrow y$ that is from iEEG spectral features to either a preictal or interictal label. The mapping is established using labeled training sets and tested to classify test sets that may have probability distribution similar to that of the training sets (see Data S1 for details).

Cost-sensitive SVMs (CSVMs) are used to handle the imbalance in number of preictal and interictal samples. Seizure events are relatively rare: There is approximately one preictal sample for every 10 interictal samples. Furthermore, we consider it more important to classify many preictal samples correctly than to have a few false positives. CSVMs allow us to set misclassification penalties of the preictal data higher than that of interictal data.

We have optimized CVSMs over two model parameters: the cost C that is a trade-off between the classification margin and misclassified or nonseparable samples and the cost-factor R between false positives (FPs, the number of interictal samples classified as preictal) and false negatives (FNs, the number of preictal samples classified as interictal). The optimal pair of C and R is selected when the average rate across the fivefold cross-validation is a
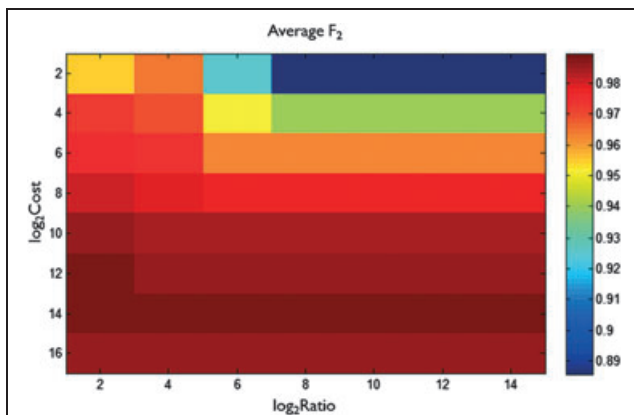
Figure 4.
Grid search with $\log_2$Cost and $\log_2$Ratio. This grid search intends to find what pair of the parameters leads to the greatest F_2 averaged over all the validation sets. The dataset used in this figure is the first training set (containing the second to fifth pre-ictal events) in Patient 17's bipolar iEEG. This example shows that the greatest $F_2 = 0.9897$ is obtained where $C = 2^{12}$ and $R = 2^2$.
Epilepsia © ILAE

maximum (See Fig. 4). We use an F_β measurement to determine the optimal pair in highly unbalanced datasets:
$F_\beta = \frac{(1+\beta^2)\cdot\mathrm{TP}}{(1+\beta^2)\cdot\mathrm{TP}+\beta^2\cdot\mathrm{FN}+\mathrm{FP}}$ (Van Rijsbergen, 1979; Li et al., 2008), where TP is the number of true positives (the number of preictal windows classified correctly). We have selected $\beta = 2$ to weight the significance of FNs more than FPs. We used the SVM-Light (Schölkopf et al., 1999) with the radial basis function kernel for classification.

We desire that SVM classification accuracy on the test data represents the actual accuracy in practice. Many previous studies trained and tested their prediction models on the same datasets (Mormann et al., 2007), resulting in overly optimistic prediction rates; their models are not free from

the *over-fitting* problem. Those prediction models, which may be excessively fit to their datasets, may result in high sensitivity and specificity in their data, but could not achieve those high rates in real conditions. To achieve less unbiased prediction rate, we use *double cross-validation* (Friedman, 1994; Cherkassky & Mulier, 2007) that ensures in-sample optimization and out-of-sample testing.

To further illustrate how we have performed double cross-validation, we will provide more details. If a patient has N seizures and I-hour–long interictal recordings, we separate the interictal datasets into N blocks (folds), each of them approximately containing continuous $\left(\frac{I}{N}\right)$-hour interictal recordings and pairing with each of 30-min preictal datasets. It is noted that interictal datasets are separated into different folds when they have a break in recordings. We randomly choose one preictal and interictal fold and reserve it for testing and use the other $(N-1)$ folds for training: N-fold cross-validation. To establish an optimal SVM classifier in training, we perform fivefold cross-validation. We randomly select 80% in the training set (the whole training set consists of $(N-1)$ folds of preictal and interictal) and establish an SVM model in the *learning* stage, and *validate* the model on the remaining 20% of the training set to check if the model is well-fit (neither over-fit nor under-fit). Once the SVM model is trained, the prediction rate is evaluated by testing the model on the fold that was reserved for testing. This process is then repeated N times and the average prediction rate is reported.

Postprocessing: removing isolated false positives

Once a test set is SVM classified, we have observed that FPs and FNs tend to be sporadic and isolated in time as compared to TPs and TNs (true negatives, the number of interictal windows classified correctly) (Netoff et al., 2009; Park et al., 2010). To eliminate these isolated FPs and FNs, we postprocess the SVM classification output using the Kalman filter (Simon, 2006; Chisci et al., 2010).
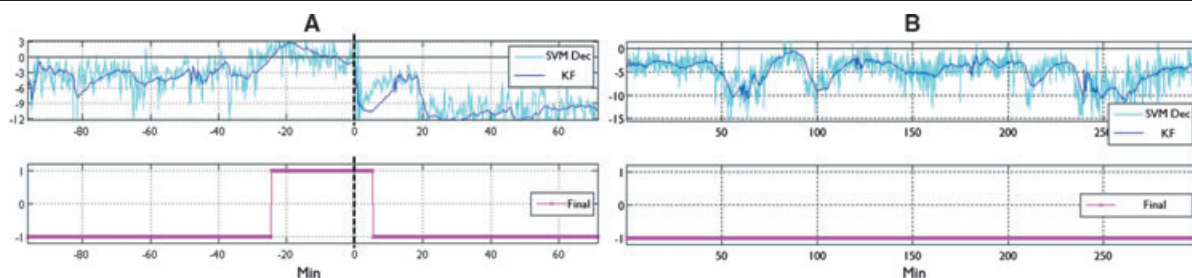

Figure 5.
Examples of SVM classification and postprocessing by the Kalman filter in testing data (**A**) with a seizure event (left, the onset is marked by black dashed lines at 0) and (**B**) with interictal recordings (right). Patient 17's bipolar iEEG recordings are used. Top panels: decision values from SVM classification (in cyan) and their Kalman-filtered outputs (in blue). If an output is >0, the sample is classified into a positive (preictal) group; otherwise, it is classified into a negative (interictal) group. Bottom panels: final outputs. Once postprocessed and classified as positive, the positive output continues in the next 30-min prediction horizon, alarming that a seizure is likely to attack within 30 min.
Epilepsia © ILAE

The Kalman filter is a statistical method that can produce estimates that tend to be close to the true values of measurements (Simon, 2006). We use the second-order discrete-time Kalman filter to reduce undesired fluctuations of the SVM classification outputs (see Data S2 for details and Fig. 5). If the Kalman filter produces a positive, then we predict that a seizure will occur within the next 30 min. If no seizure occurs within 30 min, it is considered a false positive. We choose 30 min as our prediction horizon, because we have defined preictal data as 30 min prior to a seizure in training.

KFD analysis: finding top predictive feature sets

To determine which iEEG spectral power features are key to seizure prediction, we have used the Kernel Fisher Discriminant (KFD) analysis (Mika et al., 1999; Cherkassky & Mulier, 2007). We have measured the discriminability of the spectral power features between preictal and interictal using the *F* criterion (Mika et al., 1999; Cherkassky & Mulier, 2007). Frequency bands in this analysis were not distinguished within focal or afocal electrodes. Because KFD analysis requires a balanced number of samples from each group, *F* values have been calculated on many random selections of interictal samples and averaged. Statistical Pattern Recognition Toolbox for Matlab (http://cmp.felk.c-vut.cz/cmp/software/stprtool/) has been used for KFD analysis.

RESULTS

We have tested our patient-specific binary classification algorithm for seizure prediction on iEEG of 18 patients with 80 seizure events and 433.2 h of interictal recordings in the Freiburg EEG database. To evaluate the algorithm we have measured sensitivity, the false alarm rate per hour, and the percentage of interictal recordings incorrectly classified as FPs. Sensitivity $\left(\frac{TP}{TP+FN}\right)$ measures the proportion of the preictal events in a patient classified correctly by our algorithm. The false alarm rate per hour and the percentage of interictal data that is incorrectly classified as FPs demonstrate how many false alarms the proposed algorithm would generate.

Summary results for all the patients and all the four preprocessing methods are shown in Table 1. As a baseline, prediction using spectral power features calculated from raw iEEG had total sensitivity of 93.8% (classifies 75 preictal events correctly of 80) and 0.29 false positives per hour (125 false alarm events in 433.2 h of interictal recordings), resulting in a false prediction alarm for a total of 59.2 h across all patients. Bipolar preprocessing produced a higher sensitivity of 97.5% (78 of 80) and also an improvement on the false positive rate of 0.27 per hour (118 false alarm events in 433.2 interictal hours). Time-differential preprocessing led to a significant improvement on the false positive rate of 0.20 per hour (86 false alarm events in 433.2 h), but did not improve sensitivity (a sensitivity of 92.5%, 74 of 80). Prediction based on the combined two preprocessing methods (bipolar/time-differential) demonstrated some improvement on the false positive rate to 0.23 per hour (100 false alarm events in 433.2 interictal hours) but did not improve sensitivity (93.8%, 75 of 80).

We have additionally tested other postprocessing strategies, including the *4*-of-*7* analysis (Netoff et al., 2009; Park et al., 2010) and 9-tap medial filtering, and the Kalman

Table 1. Results from seizure prediction analysis by the proposed algorithm

Pat. No.	No. of Sz	Interictal hours	Raw				Bipolar				Time-differential				Bipolar/time-differential			
			Sen%	FP/h	FP%	p-value	Sen%	FP/h	FP%	p-value	Sen%	FP/h	FP%	p-value	Sen%	FP/h	FP%	p-value
1	4	23.9	100	0.00	0.00	0.000	100	0.08	4.16	0.000	100	0.00	0.00	0.000	100	0.00	0.00	0.000
3	5	23.9	100	0.08	2.69	0.000	100	0.00	0.00	0.000	100	0.00	0.00	0.000	100	0.00	0.00	0.000
4	5	23.9	100	0.00	0.00	0.000	100	0.04	2.08	0.000	100	0.00	0.00	0.000	100	0.00	0.00	0.000
5	5	23.9	100	1.17	58.02	0.090	100	0.79	39.02	0.018	100	0.84	40.26	0.020	100	0.67	29.62	0.006
6	3	23.8	100	0.08	4.18	0.001	100	0.04	2.09	0.000	100	0.00	0.00	0.000	100	0.00	0.00	0.000
7	3	24.5	100	0.00	0.00	0.000	100	0.04	2.03	0.000	100	0.00	0.00	0.000	100	0.04	2.03	0.000
9	5	23.9	100	0.08	4.17	0.000	100	0.34	15.58	0.001	100	0.04	2.08	0.000	100	0.25	10.74	0.000
10	5	24.4	100	0.08	4.08	0.000	100	0.20	10.19	0.000	100	0.00	0.00	0.000	100	0.25	12.08	0.000
11	4	24.0	75	0.08	4.15	0.003	75	0.17	8.30	0.009	75	0.00	0.00	0.001	75	0.04	2.08	0.002
12	4	24.7	100	0.08	4.02	0.000	100	0.04	2.01	0.000	100	0.00	0.00	0.000	100	0.00	0.00	0.000
14	4	22.6	75	0.84	39.76	0.208	100	0.22	10.99	0.001	75	0.66	31.29	0.123	50	0.44	20.29	0.222
15	4	23.7	100	0.13	6.28	0.000	100	0.38	18.85	0.004	100	0.04	2.09	0.000	100	0.17	8.38	0.001
16	5	23.9	80	0.58	25.37	0.032	100	0.42	19.99	0.002	80	0.42	18.92	0.015	100	0.38	17.89	0.001
17	5	24.0	100	0.12	6.22	0.000	100	0.00	0.00	0.000	100	0.00	0.00	0.000	100	0.04	2.07	0.000
18	5	24.8	100	0.04	2.00	0.000	100	0.16	8.02	0.000	100	0.00	0.00	0.000	100	0.04	2.00	0.000
19	4	24.3	75	1.07	51.42	0.359	75	0.90	42.89	0.243	75	1.15	55.41	0.419	75	0.90	42.62	0.240
20	5	24.8	80	0.52	24.25	0.028	100	0.68	30.36	0.007	60	0.40	19.04	0.081	80	0.85	38.22	0.102
21	5	23.9	100	0.25	10.98	0.000	100	0.38	17.34	0.001	100	0.04	2.08	0.000	100	0.08	4.17	0.000
Total	80	433.2	93.8	0.29	13.67		97.5	0.27	13.01		92.5	0.20	9.45		93.8	0.23	10.69	

filter produced the best performance among them (*4-of-7* analysis resulted in 95% sensitivity and 0.41 FP/h, and 9-tap median filtering did in 95% sensitivity and 0.34 FP/h).

We have tested our results against the null hypothesis that our results could be attributed to chance (Andrzejak et al., 2003, 2009; Snyder et al., 2008). We have estimated the one-sided p-values that demonstrate how superior the sensitivity of our algorithm is to chance (Snyder et al., 2008):

$$p = 1 - \sum_{k=0}^{n-1} \binom{N}{k} S_{nc}^k (1 - S_{nc})^{N-k}, \text{ for } \frac{n}{N} \geq S_{nc},$$

where a proposed predictor correctly identifies n of N preictal events and S_{nc} is the sensitivity of the corresponding chance predictor. Prediction rates using bipolar preprocessing are significantly better than chance in 18 patients (Patient 19 excluded) at a significance cutoff of $\alpha = 0.05$ (Moore & McCabe, 2005) (See Table 1).

Two of our proposed preprocessing methods significantly enhanced the prediction rate: Bipolar preprocessing improved sensitivity and time-differential preprocessing reduced the false positive rate. Prediction by the bipolar preprocessing correctly predicted seizures in 16 patients of 18 with 100% sensitivity using a 30-min prediction window, and furthermore it missed only two seizures total (one from each of Patients 11 and 19). Time-differential preprocessing produced a significant improvement in the false positive rate. It led to a significant decrease in false alarm events to 86 from 125 in 433.2 interictal hours. Furthermore, we could perfectly predict all the preictal events generating no false alarms at all, by time-differential preprocessing, in nine patients (a total of 218.1 interictal hours), and generating only one false alarm in each of three of the other patients (in a total of 71.5 interictal hours in Patients 9, 15, and 21).

To investigate which features the SVM may be using for prediction, we have used the KFD analysis and have measured the discriminability (the F criterion) of raw or time-differential features between preictal and interictal. Discriminability of the raw features had an average F criterion of 4.58×10^{-3} over the 13 patients for whom preictal and interictal were classified well (see Table 2). Of those 13 patients, four of the patients' top discriminating features were in the gamma frequency band. Discriminability of the time-differential features was much higher (19.6×10^{-3}). Remarkably, gamma frequency bands were the most discriminating in eight patients, indicating that the time-differential preprocessing may reveal spectral changes more indicative of a preictal event.

DISCUSSION

High sensitivity of 97.5% and the low false positive trigger rate of 0.27 per hour in bipolar iEEG in 18 patients demonstrate that seizures may be predicted using the patient-specific approach of SVM classification of spectral power features. We expect this patient-specific methodology to be more successful for some patients than others; patients with repeatable and stereotypical seizures are the easiest to predict with this algorithm. By contrast, our machine learning approach may not succeed in patients whose seizures evolve rapidly over time or in patients who express several different types of seizures.

Our results may be compared directly to several other studies that have tested prediction algorithms using the same Freiburg EEG database (Aschenbrenner-Scheibe et al., 2003; Winterhalder et al., 2003; Maiwald et al., 2004; Mormann et al., 2007). We demonstrate high sensitivity and specificity for this database yet reported. Furthermore, we have used in-sample optimization and assessed the results with out-of-sample testing (a test set is never involved in training) through the use of double cross-validation experimental design. Although some studies may present higher sensitivity and/or specificity than ours (Mormann et al., 2007), their algorithms were trained and tested on the exact same datasets; therefore, the results are not directly comparable.

Another seizure prediction algorithm has been introduced recently using autoregressive coefficients and SVM classification and demonstrated a high prediction rate (Chisci et al., 2010); our algorithm is the first to be tested using double cross-validation, in that all the datasets were resampled and tested and using the prediction horizon as defined by the International Seizure Prediction Workshop (IWSP4, http://www.iwsp4.org/). The standard of the IWSP4 for prediction horizon is that once a seizure prediction has been issued it is required to be left on for the entire duration of the prediction horizon.

An important feature of our algorithm is that it analyzes linear features with a nonlinear classifier. With respect to implementing it in an implantable device, the use of linear features is attractive because they can be calculated rapidly and with low power consumption when compared to non-

Table 2. Top two discriminating raw or time-differential features

Pat. No	Raw		Time-differential	
		Top two features (F/10³)		
1	β (4.58)	γ1 (3.34)	γ4 (12.25)	γ4 (10.39)
3	γ1 (3.01)	α (2.26)	γ4 (4.88)	θ (3.80)
4	β (6.22)	γ1 (5.85)	θ (25.85)	γ3 (24.32)
6	δ (6.40)	γ1 (5.02)	θ (7.59)	γ2 (5.98)
7	γ2 (10.00)	γ3 (8.68)	γ2 (87.28)	γ3 (57.05)
9	δ (1.55)	θ (1.54)	γ4 (8.37)	γ4 (5.96)
10	β (2.04)	β (1.90)	δ (2.43)	θ (1.84)
11	θ (2.80)	γ1 (2.29)	γ4 (6.35)	γ4 (5.66)
12	θ (11.47)	α (8.19)	γ2 (80.60)	γ3 (42.39)
15	γ1 (4.20)	β (2.67)	γ1 (6.38)	γ2 (3.74)
17	θ (2.28)	γ1 (2.18)	θ (7.06)	γ1 (2.52)
18	γ1 (1.34)	γ4 (1.30)	γ4 (2.97)	γ4 (2.89)
21	δ (3.64)	α (3.64)	α (2.41)	δ (2.40)
Ave	4.58	3.76	19.57	13.00

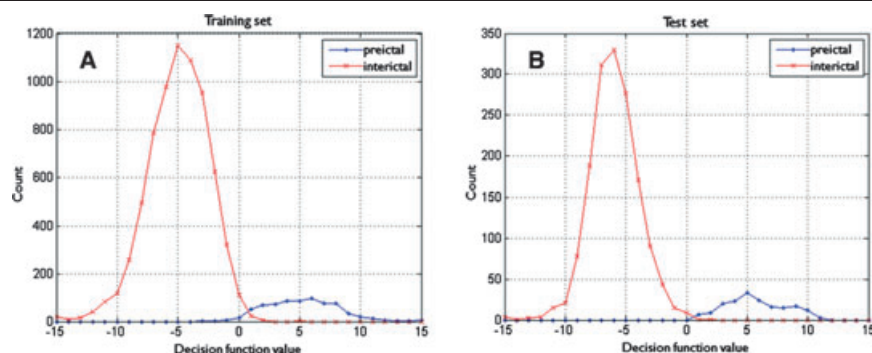The underline indicates the spectral bands in afocal electrodes.

Figure 6.
Histograms of decision function values generated from linear SVM classification. The distributions result from classification by linear SVMs (**A**) in a training set and (**B**) in the corresponding test set of Patient 17's bipolar iEEG. Blue points and red crosses represent the actual state of preictal and interictal, respectively. Positive and negative decision values represent that a sample is classified as preictal and interictal, respectively. The similar distributions of classification in training and testing by linear SVMs demonstrate that the linear SVM may classify preictal and interictal as well as the nonlinear SVM.
Epilepsia © ILAE

linear features. For classification, we used SVMs with the nonlinear kernel that may require high power consumption (Shoeb et al., 2009), but the nonlinear SVMs may be replaced by linear SVMs that consume much less power (Shoeb et al., 2009) with only a small degradation of performance (see Fig. 6).

In this study we trained and tested the algorithm after removal of artifacts. Artifacts in iEEG may negatively affect our seizure prediction algorithm and may decrease sensitivity and specificity in a real device. An additional but simple algorithm may be necessary in practice to identify the artifacts such as out-of-measurement-range or electrode disconnect in real-time.

KFD analysis of the features suggests that the SVM may predict seizures by detecting changes in gamma band power with respect to the total power. Although it has already been shown that power changes in gamma bands occur prior to a seizure and can be used for prediction (Mormann et al., 2005), our results demonstrate that using those power changes in a multivariate approach may lead to higher sensitivity and specificity. It is noted that the increase in gamma band power may be due to spontaneous iEEG spike bursts occurring more frequently rather than the emergence of intermittent rhythmic gamma waveforms (See Fig. 7). We expect that seizure prediction may improve if iEEG sampling rates are increased to detect changes in even higher frequency ranges.

The use of bipolar and/or time-differential preprocessing increased sensitivity and decreased false positive rates when compared to results from raw iEEG. Bipolar preprocessing reduces noise through common mode rejection. Time-differential preprocessing acts as a high filter and flattens the spectrum, thereby reducing the dominance of the low frequency power on the total power. The use of both bipolar
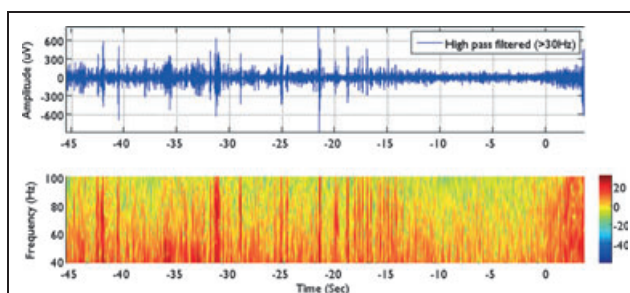

Figure 7.
High-pass filtered preictal iEEG time-trace (>30 Hz, top) and its original signal's spectrogram (bottom).
Epilepsia © ILAE

and time-differential methods provided low false-positive rate without decreasing sensitivity.

Seizure prediction may be further improved by optimal selection of preictal data for training the SVM algorithm. In this study, we have operationally defined preictal as any data that occurred 30 min prior to a seizure. Our algorithm successfully distinguished iEEG in 30 min prior to the seizures from interictal recordings. This indicates that a heightened seizure state could be detected but it could not be determined if that change of states occurs at a certain finite time prior to the seizures. When our classifier was applied to the iEEG recordings that contain ones immediately preceding 30-min preictal, it could distinguish 30-min preictal periods only in some cases (10.3%, 8 of 78). In the strict sense, our seizure prediction algorithm "anticipated" seizures (Mormann et al., 2007), where an algorithm can successfully identify a seizure occurrence in advance but may not be able to specify a time when the event will occur. It is possible that many of the patients might have actual preictal

periods longer than 30 min or that seizure states might change prior to the onset of the recordings. Better prediction may be achieved by selecting correct preictal periods for each patient. It is noted that interictal iEEG recordings in the Freiburg database were separated from seizure events generally by several days and therefore it cannot be ruled out that our classifier might only be detecting changes in the iEEG due to withdrawal of antiepileptic drugs.

Prediction may be also improved by adding and testing more features, including other univariate features such as autoregressive coefficients (Chisci et al., 2010) or bivariate features such as cross-correlation or wavelet synchrony (Mirowski et al., 2009). Selecting discriminating features may also enhance the classification rate (Guyon & Elisseeff, 2003; Saeys et al., 2007) and furthermore improve power consumption when it comes to a real device. To reduce false positive time, a criterion may be established to turn off (false) alarms, as new windows of data are analyzed: for example, a Kalman-filtered output is marked as an (false) alarm but it goes off immediately. Under this criterion, which may be against the IWSP4's standard, we could have achieved a total of 24.8 h false positives (5.72%) in 433.2 h bipolar interictal recordings. Support Vector Data Description that is modified one-class SVM with cost-sensitive learning may provide higher sensitivity and specificity in seizure prediction than cost-sensitive SVMs (Tax & Duin, 2004). Lastly, establishing a continuous variable that indicates the likelihood of an impending seizure in the near future can be an alternative approach to seizure prediction.

Acknowledgments

Disclosures

Yun Park, Lan Luo, Keshab K. Parhi, and Theoden Netoff have no conflicts of interest to report. 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

Alpaydin E. (2004) *Introduction to machine learning.* The MIT Press, Cambridge.

Andrzejak R, Mormann F, Kreuz T, Rieke C, Kraskov A, Elger C, Lehnertz K. (2003) Testing the null hypothesis of the nonexistence of a preseizure state. *Phys Rev E Stat Nonlin Soft Matter Phys* 67:10901.

Andrzejak R, Chicharro D, Elger C, Mormann F. (2009) Seizure prediction: any better than chance? *Clin Neurophysiol* 120:1465–1478.

Aschenbrenner-Scheibe R, Maiwald T, Winterhalder M, Voss H, Timmer J, Schulze-Bonhage A. (2003) How well can epileptic seizures be predicted? An evaluation of a nonlinear method. *Brain* 126:2616.

Bishop C. (2006) *Pattern recognition and machine learning.* Springer, New York.

Cherkassky V, Mulier F. (2007) *Learning from data: concepts, theory, and methods.* Wiley-Interscience, Hoboken.

Chisci L, Mavino A, Perferi G, Sciandrone M, Anile C, Colicchio G, Fuggetta F. (2010) Real-time epileptic seizure prediction using AR models and support vector machines. *IEEE Trans Biomed Eng* 57:1124–1132.

Feldwisch-Drentrup H, Schelter B, Jachan M, Nawrath J, Timmer J, Schulze-Bonhage A. (2010) Joining the benefits: combining epileptic seizure prediction methods. *Epilepsia* 51:1598–1606.

Fisher R, Salanova V, Witt T, Worth R, Henry T, Gross R, Oommen K, Osorio I, Nazzaro J, Labar D. (2010) Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy. *Epilepsia* 51:899–908.

Friedman JH. (1994) An overview of predictive learning and function approximation. In Cherkassky V, Friedman JH, Wechsler H (Eds) *From statistics to neural networks: Theory and pattern recognition applications (NATO ASI Series/Computer and Systems Sciences).* Springer, Berlin, pp. 1–61.

Guyon I, Elisseeff A. (2003) An introduction to variable and feature selection. *JMLR* 3:1157–1182.

Harrison M, Osorio I, Frei M, Asuri S, Lai Y. (2005) Correlation dimension and integral do not predict epileptic seizures. *Chaos* 15:033106.

Hjorth B. (1970) EEG analysis based on time domain properties. *Electroencephalogr Clin Neurophysiol* 29:306.

Li X, Wang Y, Acero A. (2008) Learning query intent from regularized click graphs. *Proc. ACM Int. Conf. Inf. Knowl. Manag.* ACM Press, Singapore, pp. 339–346.

Litt B, Echauz J. (2002) Prediction of epileptic seizures. *Lancet Neurol* 1:22–30.

Maiwald T, Winterhalder M, Aschenbrenner-Scheibe R, Voss H, Schulze-Bonhage A, Timmer J. (2004) Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physica D* 194:357–368.

Martinerie J, Adam C, Le Van Quyen M, Baulac M, Clemenceau S, Renault B, Varela F. (1998) Epileptic seizures can be anticipated by non-linear analysis. *Nat Med* 4:1173–1176.

Mika S, Ratsch G, Weston J, Scholkopf B, Muller K. (1999) Fisher discriminant analysis with kernels. *Neural networks for signal processing IX.* IEEE, Madison, WI, U.S.A., pp. 41–48.

Mirowski P, Madhavan D, LeCun Y, Kuzniecky R. (2009) Classification of patterns of EEG synchronization for seizure prediction. *Clin Neurophysiol* 120:1927–1940.

Moore D, McCabe G. (2005) *Introduction to the practice of statistics.* W.H. Freeman, New York.

Mormann F, Kreuz T, Rieke C, Andrzejak R, Kraskov A, David P, Elger C, Lehnertz K. (2005) On the predictability of epileptic seizures. *Clin Neurophysiol* 116:569–587.

Mormann F, Andrzejak R, Elger C, Lehnertz K. (2007) Seizure prediction: the long and winding road. *Brain* 130:314.

Netoff T, Park Y, Parhi K. (2009) Seizure prediction using cost-sensitive support vector machine. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* IEEE, Minneapolis, MN, U.S.A., pp. 3322–3325.

Nunez P, Srinivasan R. (2006) *Electric fields of the brain: the neurophysics of EEG.* Oxford University Press, New York.

Osorio I, Frei M, Wilkinson S. (1998) Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia* 39:615–627.

Park Y, Neoff T, Parhi K. (2010) Seizure prediction with spectral power of time/space-differential EEG signals using cost-sensitive support vector machine. *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, Dallas, TX, U.S.A., pp. 5450–5453.

Rothman S, Smyth M, Yang X, Peterson G. (2005) Focal cooling for epilepsy: an alternative therapy that might actually work. *Epilepsy Behav* 7:214–221.

Saeys Y, Inza I, Larrañaga P. (2007) A review of feature selection techniques in bioinformatics. *Bioinformatics* 23:2507–2517.

Schölkopf B, Burges C, Smola A. (1999) *Advances in kernel methods: support vector learning*. The MIT press, Cambridge, MA.

Shoeb A, Carlson D, Panken E, Timothy D. (2009) A micro support vector machine based seizure detection architecture for embedded medical devices. *Conf. Proc. IEEE Eng. Med. Biol. Soc*. IEEE, Minneapolis, MN, U.S.A., pp. 4202–4205.

Simon D. (2006) *Optimal state estimation: Kalman, H [infinity] and non-linear approaches*. John Wiley and Sons, Hoboken, NJ.

Snyder D, Echauz J, Grimes D, Litt B. (2008) The statistics of a practical seizure warning system. *J Neural Eng* 5:392.

Tax D, Duin R. (2004) Support vector data description. *Mach Learn* 54:45–66.

Van Rijsbergen C. (1979) *Information retrieval*. Butterworth-Heinemann, London.

Vapnik V. (2000) *The nature of statistical learning theory*. Springer, New York.

Winterhalder M, Maiwald T, Voss H, Aschenbrenner-Scheibe R, Timmer J, Schulze-Bonhage A. (2003) The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. *Epilepsy Behav* 4:318–325.

Yang X, Schmidt B, Rode D, Rothman S. (2009) Optical suppression of experimental seizures in rat brain slices. *Epilepsia* 51:127–135.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. SVM classification using double cross-validation.

Data S2. Postprocessing by the Kalman filter.

Figure S1. Support vector machines (Cherkassky & Mulier, 2007). In this figure, circles and squares are classified in a high-dimensional space by the decision boundary of the optimal hyperplane (indicated as a solid line), maximizing the margin (indicated as the distance between the two dashed lines) that are formed by the support vectors (in gray).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.