Distributed Kriging-Bootstrapped DNN Model for Real-Time Seizure Detection from EEG Signals

I. L. Olokodana¹, S. P. Mohanty², and E. Kougianos³
Smart Electronic Systems Laboratory (SESL) ¹,²,³
University of North Texas, TX, USA.¹,²,³

Email: IbrahimOlokodana@my.unt.edu¹, saraju.mohanty@unt.edu², elias.kougianos@unt.edu³
Outline of the talk

- Epileptic Seizures
- Kriging Methods
- Novel Contributions
- Brain as a Spatial Object
- Proposed Seizure Detection Models
- Experimental Results
- Conclusion & Future Research
What is Seizure?

- A seizure is an abnormal activity in the nervous system which causes its sufferers to lose consciousness and control.
Why Seizure Detection?

- 4th most common neurological disease in the world.
- Potential incidence rate is about 10% of world population.
- Mortality rate of about 44% compared to the general population with 12.2% mortality rate.
Types of Seizure

- Epileptic Seizures
  - Generalized Seizures
    - Generalized Tonic-Clonic
  - Focal or Partial Seizures
    - Absence
    - Myoclonic
  - Epileptic Spasm
    - Atonic
      - Typical
      - Atypical
Electroencephalogram (EEG) Signals

 Characteristics
 - High Complexity
 - Low Intensity
 - Frequency: 0.5–30Hz

Seizure EEG Signal
What are the Research Problems?

- Automatic Seizure Detection.
- Seizure Detection Latency.
- Extended Training Time.
- Patient-Specific Seizure Detection.
- Mobile and Portable Seizure Detection.
- Seizure Crisis Intervention Mechanism.
- Seamless, consistently accurate seizure detection.
- Low-power seizure detection system.
What are the Challenges?

- Collection of custom dataset.
- Testing directly on human or animal subjects.
- Noise due to artifacts and environmental factors.
## Related Research in Seizure Detection – EEG/ML

<table>
<thead>
<tr>
<th>Reference</th>
<th>Dataset</th>
<th>Processing</th>
<th>Classifier Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alejandro et al (2017)</td>
<td>CHB-MIT</td>
<td>Time Analysis</td>
<td>Thresholds</td>
</tr>
<tr>
<td>Zhou et al. (2012)</td>
<td>Freiburg Database</td>
<td>Wavelet</td>
<td>Bayesian Method</td>
</tr>
<tr>
<td>Acharya et al. (2012)</td>
<td>Self-recorded</td>
<td>Freq. Analysis</td>
<td>SVM, KNN</td>
</tr>
<tr>
<td>Khan et al. (2012)</td>
<td>Self-recorded</td>
<td>Wavelet</td>
<td>LDA</td>
</tr>
<tr>
<td>Runarsson et al (2005)</td>
<td>Self-recorded</td>
<td>Time Analysis</td>
<td>SVM</td>
</tr>
<tr>
<td>Rezvan et al. (2017)</td>
<td>Bonn Dataset</td>
<td>Wavelet</td>
<td>MLP</td>
</tr>
<tr>
<td>Mursalin et al. (2017)</td>
<td>Bonn Dataset</td>
<td>Time Analysis</td>
<td>Random Forest</td>
</tr>
<tr>
<td>Guo et al. (2010)</td>
<td>Bonn Dataset</td>
<td>Wavelet</td>
<td>ANN</td>
</tr>
<tr>
<td>Mitra et al. (2009)</td>
<td>Texas’ Children</td>
<td>Freq. Analysis</td>
<td>ANN</td>
</tr>
<tr>
<td>Zandi et al. (2010)</td>
<td>Vancouver GH</td>
<td>Wavelet</td>
<td>Thresholds</td>
</tr>
</tbody>
</table>
Related Research in Seizure Detection – Non-EEG


IBM’s Implantable Seizure Detector

- The TrueNorth chip is postage stamp-sized and consumes over 1,000 times less power than a conventional processor of similar size.

Source: http://uberveillance.squarespace.com/?category=health_care
Consumer Electronics for Seizure Detection

- **Embrace2**: Smart-band which uses machine learning to detect convulsive Seizures and notifies caregivers.

- **Medical grade smart watch**: It detects generalized clonic-tonic Seizures and notifies physicians.

Source: https://spectrum.ieee.org/the-human-os/biomedical/diagnostics/this-seizuredetecting-smartwatch-could-save-your-life

Source: https://www.empatica.com/embrace2/
What are the Drawbacks of Existing Works?

- High seizure detection latency.
- Not suitable for real time IoMT deployment.
- Intervention mechanism after detection is lacking.
Research Question and Hypothesis

Is it possible to achieve a further reduction in training time by distributing the Kriging computation across different cores without affecting the overall performance of the seizure detection model?
Novel Contributions

➢ Novel hierarchical and distributed Kriging-Bootstrapped Deep Neural Network (DNN) models for seizure detection.

➢ Achievement of a seizure detection latency of less than 1 sec.

➢ A novel single-channel seizure detection.

➢ 91% reduction in training time & performance improvement by at least 2.5%.
Kriging

- It is a Gaussian process dependent on mean and co-variances of data points.
Why Kriging?

- The modeling of the brain as a geo-spatial map.
- Good performance on small datasets.
- Estimation variance.
- Few hyperparameters.
Existing Applications of Kriging

- Seismic intensity analysis (De Rubeis et al., 2005).
- Hydrology and well selection (Virdee et al., 1984).
- Geodesy and geology (Reguzzoni et al., 2005).
- Structural reliability (Kaymaz et al., 2005).
- Mixed signal design optimization (Mohanty et al., 2015).
- Cellular network optimization (Braham et al., 2014).
Brain as a Spatial Map


Kriging Estimates

Given the following set of observations $x_1, x_2, ..., x_n$ as inputs, and $y(x_1), y(x_2), ..., y(x_n)$ as outputs, the input-output relationship based on Kriging is given by:

1. $y(x_i) = \mu + Z(x_i)$

A residual equation can then be written as:

2. $y(x_0) - \mu_z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) - \mu_z(x_i)$

Hence, a linear estimation for an unknown is formulated as follows: 3.

$$y(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) + (1 - \sum_{i=1}^{n} \lambda_i) \mu_z$$

If we let $y = Z^*$ and represent a vector of residuals with $R$, then the residual equation in Eqn. 2. becomes:

4. $R^*(x_0) = \sum_{i=1}^{n} \lambda_i R(x_i)$

The estimation variance of Kriging’s prediction is given by:

5. $\sigma^2_{est.} = E\{[R^*(x_0) - R(x_0)]^2\}$

Expanding the equation gives rise to: 6.

$\sigma^2_{est.} = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C(x_i, x_j) - 2 \sum_{i=1}^{n} \lambda_i C(x_0, x_i) + C(0)$

The partial derivative of Eqn. 6. above with respect to $\lambda_i$ results in:

7. $\frac{\partial \sigma^2_{est.}}{\partial \lambda_i} = \sum_{j=1}^{n} \lambda_j C(x_i, x_i) - 2 C(x_0, x_i)$

Where $i = 1, 2, 3, ..., n$. By setting Eqn. 7. to zero, we have a system of $n$ equations and $n$ unknown weights as follows:

8. $\sum_{j=1}^{n} \lambda_j C(x_i, x_i) = 2 C(x_0, x_i)$

The weights $\lambda_i$ can then finally be obtained by solving Eqn. 8. Hence, the kriging estimate $y(x_0)$ can be obtained from Eqn 3.

The covariance function is given by:

9. $C_x(h) = \sigma^2_x(h) - \gamma(h)$
The Bootstrap

- A Bootstrap helps in pulling on a boot.
- It means solving a problem without external resources

Source: http://www.lemen.com/dictionary-b.html#bootstrap
Bootstrap Sampling

Draw and calculate Statistic B times

1. Draw $x_1, \ldots, x_n$ from $P$
   Compute $M_1 = g(x_1, \ldots, x_n)$

2. Draw $x_1, \ldots, x_n$ from $P$
   Compute $M_2 = g(x_1, \ldots, x_n)$

   ... 

B. Draw $x_1, \ldots, x_n$ from $P$
   Compute $M_B = g(x_1, \ldots, x_n)$

Get B Statistic

Summarize

Mean

$$\bar{m} = \frac{1}{B} \sum_{j=1}^{B} M_j$$

Variance

$$s^2 = \frac{1}{B} \sum_{j=1}^{B} (M_j)^2 - \left( \frac{1}{B} \sum_{j=1}^{B} M_j \right)^2$$

https://towardsdatascience.com/an-introduction-to-the-bootstrap-method-58bcb51b4d60
Bootstrapped Kriging

- Initial Sample Field
  - Sample Size = 7
- New Sample Field
  - Bootstrap Size = 13
Deep Neural Network

- The neural network operates by minimizing the cost function.

- The cost function is given by the following expression:

\[
J(\omega, b) = -\frac{1}{n} \sum_{i=1}^{n} [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]
\]

\[
\omega := \omega - \alpha \frac{\partial J(\omega, b)}{\partial \omega}, \quad b := b - \alpha \frac{\partial J(\omega, b)}{\partial \omega}
\]

- Where \(\omega\) refers to weights, \(b\) is bias, \(\alpha\) is learning rate, \(n\) is number of samples. \(y_i\) and \(\hat{y}_i\) are true are predicted values respectively.

\[
\hat{y} = f(\omega^T X + b)
\]
Proposed Distributed Kriging Model

[Graph showing distributed cores for Kriging model]
Proposed Distributed Kriging-Bootstrapped DNN Model

- Extracted Features:
  - HJ – Hjorth Complexity
  - EN – Signal Entropy
  - FD – Fractal Dimension

- Distributed Kriging Model

- L-Layer Deep Neural Network (DNN) Model
Computational Analysis of Distributed Kriging

\[ \sum_{j=1}^{n} \lambda_j C(x_i, x_j) = C(x_0, x_i) \]

\[
\begin{bmatrix}
C(x_1, x_1) & C(x_1, x_2) & C(x_1, x_3) \\
C(x_2, x_1) & C(x_2, x_2) & C(x_2, x_3) \\
C(x_3, x_1) & C(x_3, x_2) & C(x_3, x_3)
\end{bmatrix}
= \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}
\begin{bmatrix}
C(x_0, x_1) \\
C(x_0, x_2) \\
C(x_0, x_3)
\end{bmatrix}
\]

\[
\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix}
C(x_1, x_1) & C(x_1, x_2) & C(x_1, x_3) \\
C(x_2, x_1) & C(x_2, x_2) & C(x_2, x_3) \\
C(x_3, x_1) & C(x_3, x_2) & C(x_3, x_3)
\end{bmatrix}^{-1}
\begin{bmatrix}
C(x_0, x_1) \\
C(x_0, x_2) \\
C(x_0, x_3)
\end{bmatrix}
\]
Proposed Fast & Accurate Real-time Seizure Detection Model
Training or Learning Process

Input \( n \) → Quasi-Output \( \frac{B}{N} \) → \( n+B \) → Final-Output \( S_{Status} \)

\[ \text{Kriging} \]
BONN DATASET (DATASET A)

- Consists of 5 sets from A-E. Sets A&B are healthy signals, C&D are inter-ictal signals while E is the only set with ictal signals. Each of the sets comprises 100 EEG segments which were collected with a 128-channel EEG system sampled at 173.61 Hz.
The EEG signals were collected from 22 epileptic patients of CHB using a 23-channel EEG, sampled at 256Hz and labeled according to the subjects as chb01 to chb23.

The dataset consists of a total of 916 hours of continuous EEG recordings across all 22 subjects.

Data from 5 patients were used in this work. They include chb01, chb03, chb05, chb07 and chb09.
Extracted Features

- Signal Entropy
- Fractal Dimension
- Signal Power
- Standard Deviation
- Singular Value Decomposition Entropy
- Maximum Fractal Length
- Hjorth Parameters
- Hurst Exponent
- Lyapunov Exponent etc.
Dataset and Features

The features extracted are: Hjorth Complexity (HJC), SVD Entropy (ENT), Petrosian Fractal Dimension (PFD)

\[ HJC = \frac{\text{Mobility} \left( \frac{dy(t)}{dt} \right)}{\text{Mobility} (y(t))} \]

where

\[ \text{Mobility} = \sqrt{\frac{\text{var} \left( \frac{dy(t)}{dt} \right)}{\text{var} (y(t))}} \]

\[ ENT = \sum_{i=1}^{M} \bar{\sigma}_i \log_2 (\bar{\sigma}_i) \]

\[ PFD = \frac{\log_e (n)}{\log_e (n) + \log_e \left( \frac{n}{n + 0.4N_\delta} \right)} \]
Experimental Results: Dataset A


<table>
<thead>
<tr>
<th>DNN Specifications</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of Layers</td>
<td>4</td>
</tr>
<tr>
<td>Hidden Units</td>
<td>5, 5, 3</td>
</tr>
<tr>
<td>Hidden L. Activation</td>
<td>Rectified Linear Unit (ReLU)</td>
</tr>
<tr>
<td>Output L. Activation</td>
<td>Sigmoid Function</td>
</tr>
<tr>
<td>Initialization Method</td>
<td>Xavier Initialization</td>
</tr>
<tr>
<td>Optimization Method</td>
<td>Adaptive Momentum</td>
</tr>
</tbody>
</table>

- Distributed kriging-bootstrapped DNN model performance with 10,000 samples using Dataset A.

<table>
<thead>
<tr>
<th>Count</th>
<th>Training Accuracy</th>
<th>Testing Accuracy</th>
<th>Training Epochs</th>
<th>Training Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>99.14%</td>
<td>97.50%</td>
<td>500</td>
<td>12.80s</td>
</tr>
<tr>
<td>2</td>
<td>99.76%</td>
<td>100.00%</td>
<td>800</td>
<td>13.46s</td>
</tr>
<tr>
<td>3</td>
<td>99.84%</td>
<td>100.00%</td>
<td>1000</td>
<td>13.75s</td>
</tr>
<tr>
<td>4</td>
<td>99.92%</td>
<td>100.00%</td>
<td>1500</td>
<td>15.56s</td>
</tr>
<tr>
<td>5</td>
<td>99.92%</td>
<td>100.00%</td>
<td>10000</td>
<td>52.72s</td>
</tr>
</tbody>
</table>
Experimental Results: Dataset A

- Training Time reduced by 91%

<table>
<thead>
<tr>
<th>Models</th>
<th>DNN</th>
<th>Ordinary Kriging</th>
<th>Kriging DNN</th>
<th>Distributed Kriging DNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tr. Data Size</td>
<td>10000</td>
<td>2000</td>
<td>10000</td>
<td>10000</td>
</tr>
<tr>
<td>Tr. Epochs</td>
<td>45000</td>
<td>NA</td>
<td>1500</td>
<td>1500</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.00001</td>
<td>NA</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Training Acc.</td>
<td>99.99%</td>
<td>100.00%</td>
<td>99.92%</td>
<td>99.92%</td>
</tr>
<tr>
<td>Testing Acc.</td>
<td>97.50%</td>
<td>99.78%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Training Time</td>
<td>173.57s</td>
<td>72.24s</td>
<td>43.83s</td>
<td>15.56s</td>
</tr>
</tbody>
</table>
### Experimental Results: Dataset A

<table>
<thead>
<tr>
<th>Models</th>
<th>Detection Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNN</td>
<td>0.80s</td>
</tr>
<tr>
<td>Ordinary Kriging</td>
<td>0.86s</td>
</tr>
<tr>
<td>Krig-DNN</td>
<td>0.80s</td>
</tr>
<tr>
<td>Dist-Krig-DNN</td>
<td>0.80s</td>
</tr>
</tbody>
</table>

![Graphs showing experimental results](graph1.png)
Experimental Results: Dataset B

- Comparing the performances of single and multi-channel Models using Dataset B.

<table>
<thead>
<tr>
<th>Models</th>
<th>Channel Type</th>
<th>No of Channels</th>
<th>Training Accuracy</th>
<th>Testing Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kriging</td>
<td>Single</td>
<td>1</td>
<td>68.00%</td>
<td>59.00%</td>
</tr>
<tr>
<td>Kriging</td>
<td>Multiple</td>
<td>23</td>
<td>99.70%</td>
<td>89.00%</td>
</tr>
<tr>
<td>Dist-Krig-DNN</td>
<td>Single</td>
<td>1</td>
<td>100.00%</td>
<td>98.53%</td>
</tr>
</tbody>
</table>

- Comparing best performances for basic DNN and distributed Kriging-bootstrapped DNN models using Dataset B.

<table>
<thead>
<tr>
<th>Models</th>
<th>Training Accuracy</th>
<th>Testing Accuracy</th>
<th>Training Epochs</th>
<th>Training Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNN</td>
<td>99.97%</td>
<td>98.53%</td>
<td>10000</td>
<td>42.03s</td>
</tr>
<tr>
<td>Dist-Krig-DNN</td>
<td>100.00%</td>
<td>98.53%</td>
<td>500</td>
<td>7.05s</td>
</tr>
</tbody>
</table>

- The result shows that the training time of our proposed model is reduced by 83% compared to the baseline DNN and also trains in 20 times less training epochs.
Experimental Results: Dataset B

- Ordinary Kriging: 80.7%
- Distributed Kriging: 51.1%
- DNN: 19.3%
- DNN: 48.9%


The accuracy of our proposed single-channel seizure detection model compares well with multichannel models in existing works.

<table>
<thead>
<tr>
<th>Published Works</th>
<th>Dataset</th>
<th>Classification Algorithm</th>
<th>Channel Type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoo et al., 2013</td>
<td>CHB-MIT</td>
<td>SVM</td>
<td>Multichannel</td>
<td>84.4%</td>
</tr>
<tr>
<td>Sabrina et al. 2016</td>
<td>CHB-MIT</td>
<td>Clustering</td>
<td>Multichannel</td>
<td>98.84%</td>
</tr>
<tr>
<td>Mengni et al. 2018</td>
<td>CHB-MIT</td>
<td>CNN</td>
<td>Multichannel</td>
<td>97.5%</td>
</tr>
<tr>
<td>Ye Yuan et al., 2018</td>
<td>CHB-MIT</td>
<td>WT-CtxFusion</td>
<td>Multichannel</td>
<td>95.71%</td>
</tr>
<tr>
<td>Chulkyun, et al., 2018</td>
<td>CHB-MIT</td>
<td>CNN</td>
<td>Multichannel</td>
<td>85.6%</td>
</tr>
<tr>
<td>Current Paper</td>
<td>CHB-MIT</td>
<td>Dist-Kriging-Bootstrapped DNN</td>
<td>Single Channel</td>
<td>98.53%</td>
</tr>
</tbody>
</table>
Conclusions

- The detection of seizure onset takes place in near real time with an average detection latency of 0.80 second which is better than previous models in the literature.

- A downward spiral in training time up to about 91% reduction compared to a baseline model was achieved with a novel single-channel seizure detection model which showed a better performance than existing multichannel models.
Future Research

- In future, we will investigate seizure prediction, which means having prior knowledge that a seizure will occur before it actually does.
- Another future research is to have unified systems that detects seizure before it happens, and then injects drug or performs other control measures right after that.
- We also intend to add security and privacy features to the overall system as it is IoMT-enabled and always connected to Internet.
- We will also use more sophisticated and power-efficient edge devices such as IBM’s neurosynaptic hardware in validating our models.
References


THANK YOU