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

I. L. Olokodana[1], S. P. Mohanty[2], and E. Kougianos[3]

Smart Electronic Systems Laboratory (SESL) [1,2,3]

University of North Texas, TX , USA.[1,2,3]

Email: IbrahimOlokodana@my.unt.edu[1], saraju.mohanty@unt.edu[2], elias.kougianos@unt.edu[3]

Outline of the talk

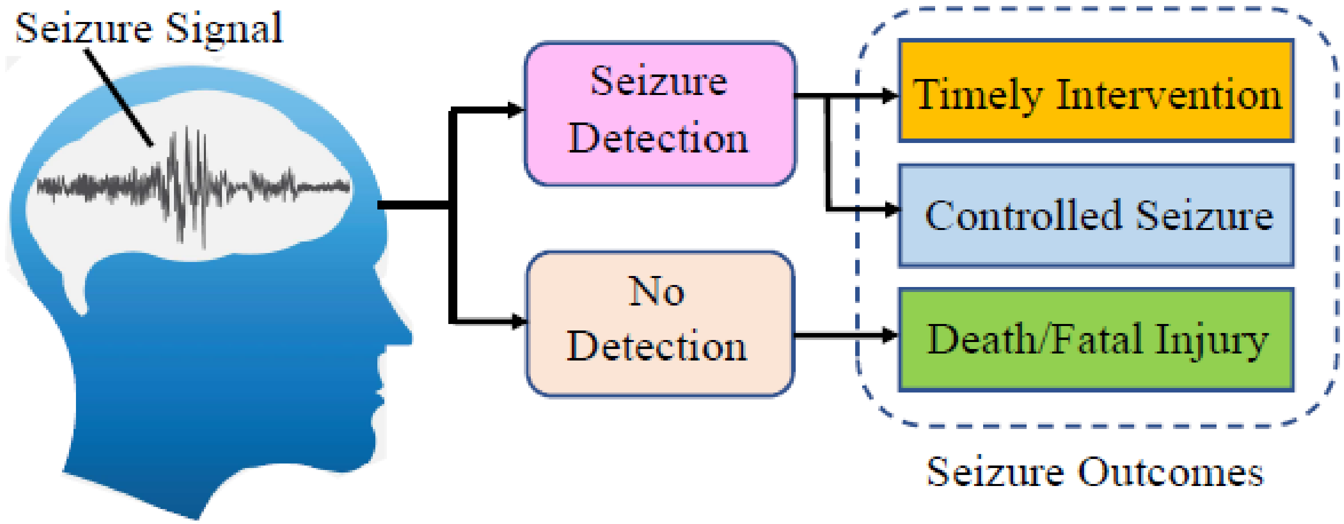
- Epileptic Seizures
- Kriging Methods
- Novel Contributions
- Brain as a Spatial Object
- Proposed Seizure Detection Models
- Experimental Results
- Conclusion & Future Research

What is Seizure?

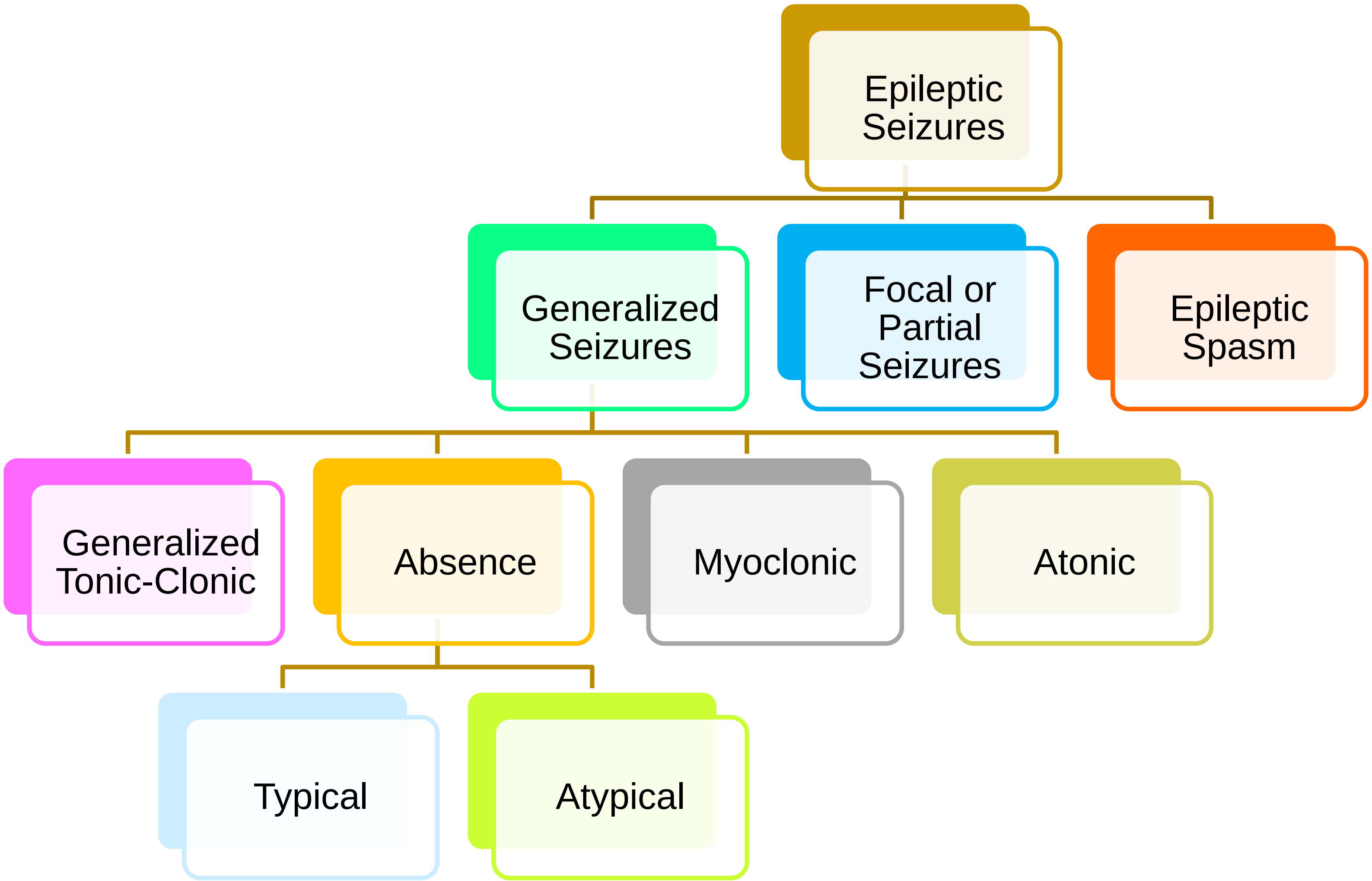
- A seizure is an abnormal activity in the nervous system which causes its sufferers to lose consciousness and control.

Why Seizure Detection?

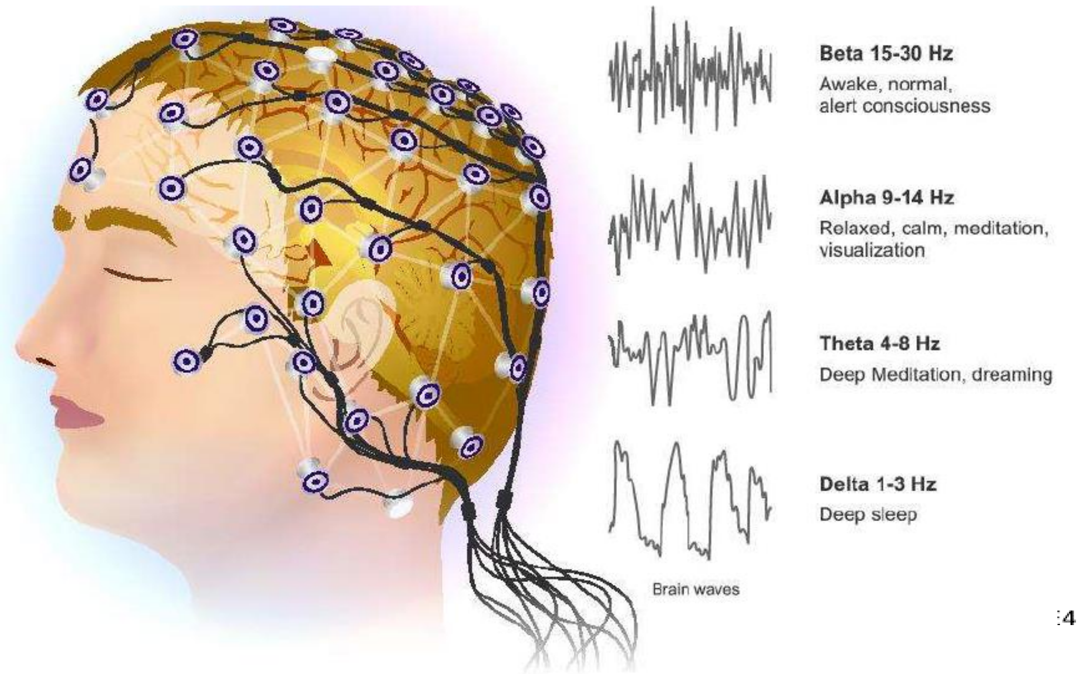
- 4th most common neurological disease in the world.
- Potential incidence rate is about 10% of world population.
- Mortality rate of about 44% compared to the general population with 12.2% mortality rate.



Types of Seizure



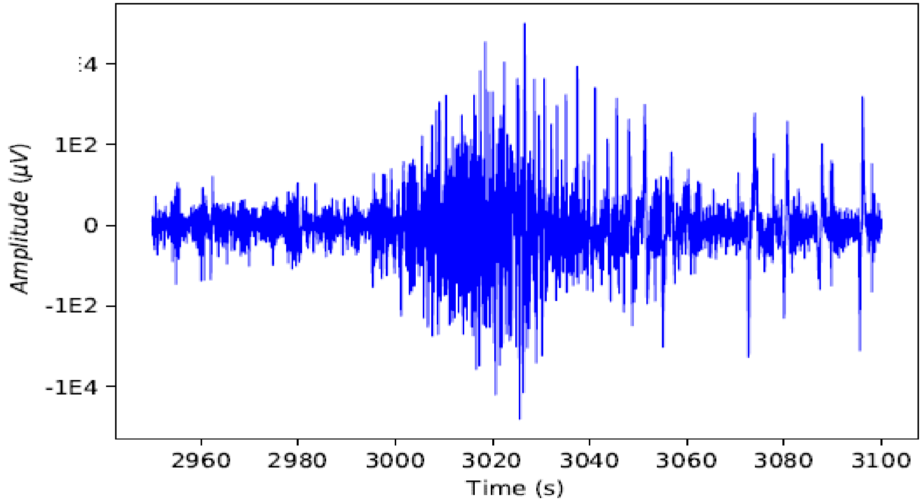
Electroencephalogram (EEG) Signals



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

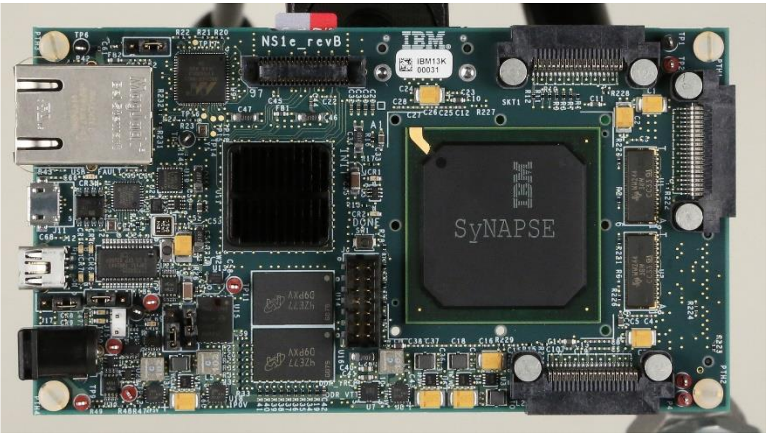
Reference	Dataset	Processing	Classifier Type
Alejandro et al (2017)	CHB-MIT	Time Analysis	Thresholds
Zhou et al. (2012)	Freiburg Database	Wavelet	Bayesian Method
Acharya et al. (2012)	Self-recorded	Freq. Analysis	SVM, KNN
Khan et al. (2012)	Self-recorded	Wavelet	LDA
Runarsson et al (2005)	Self-recorded	Time Analysis	SVM
Rezvan et al. (2017)	Bonn Dataset	Wavelet	MLP
Mursalin et al. (2017)	Bonn Dataset	Time Analysis	Random Forest
Guo et al. (2010)	Bonn Dataset	Wavelet	ANN
Mitra et al. (2009)	Texas' Children	Freq. Analysis	ANN
Zandi et al. (2010)	Vancouver GH	Wavelet	Thresholds

Related Research in Seizure Detection – Non-EEG

- A. Marquez, M. Dunn, J. Ciriaco, and F. Farahmand, “iSeiz: A low-cost real-time seizure detection system utilizing cloud computing,” in Proc. IEEE Glob. Hum. Tech. Conf., 2017, pp. 1–7.
- P. M. Vergara, E. de la Cal, J. R. Villar, V. M. Gonz´alez, and J. Sedano, “An IoT platform for epilepsy monitoring and supervising,” J. Sensors, vol. 2017, July 2017.
- Pavei, J., Heinzen, R. G., Novakova, B., Walz, R., Serra, A. J., Reuber, M., ... & Marques, J. L. (2017). Early seizure detection based on cardiac autonomic regulation dynamics. Frontiers in physiology, 8, 765.

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

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

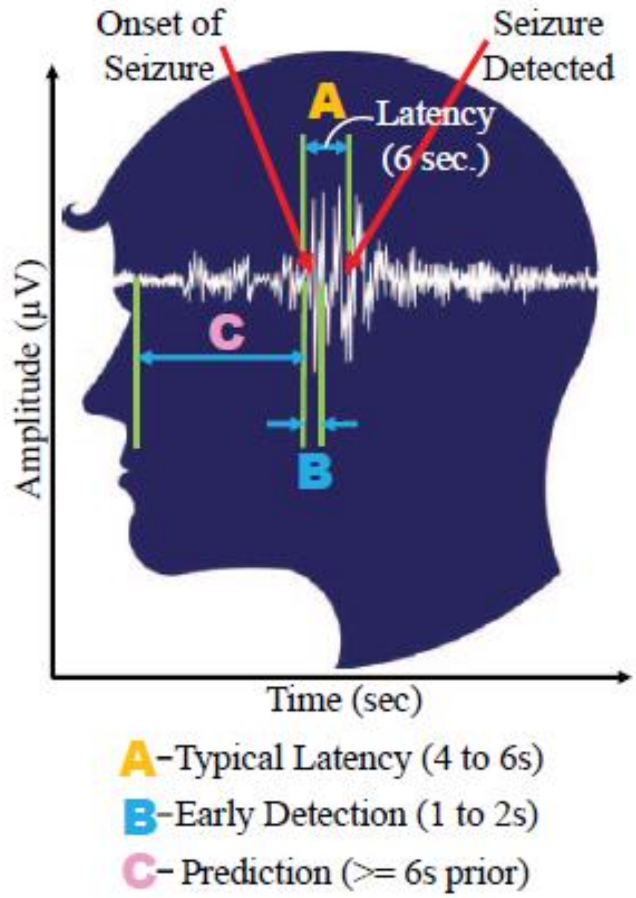
- Embrace2: Smart-band which uses machine learning to detect convulsive Seizures and notifies caregivers.

Source: https://www.empatica.com/embrace2/

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

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

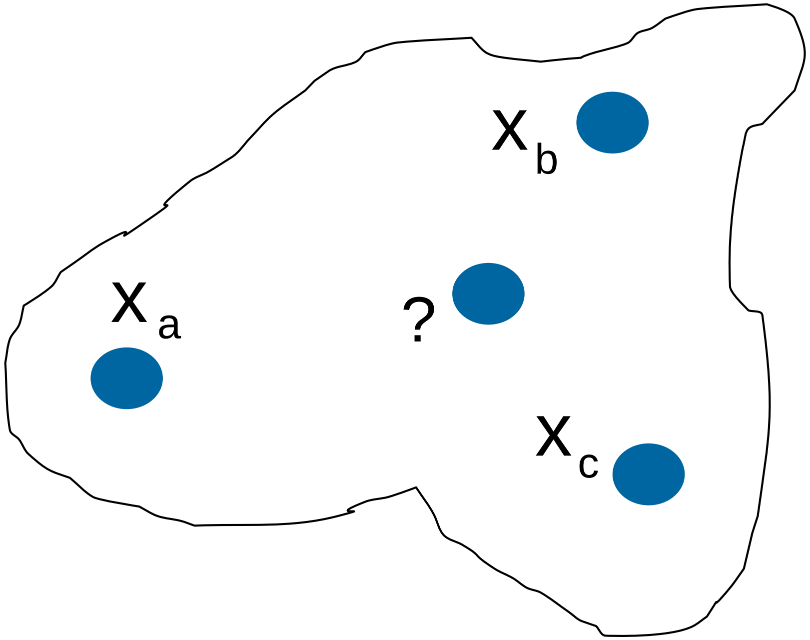
- Is it possible to achieve a further reduction in training time by distributing the Kriging computation across different cores without affecting the overall performance of the seizure detection model?

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



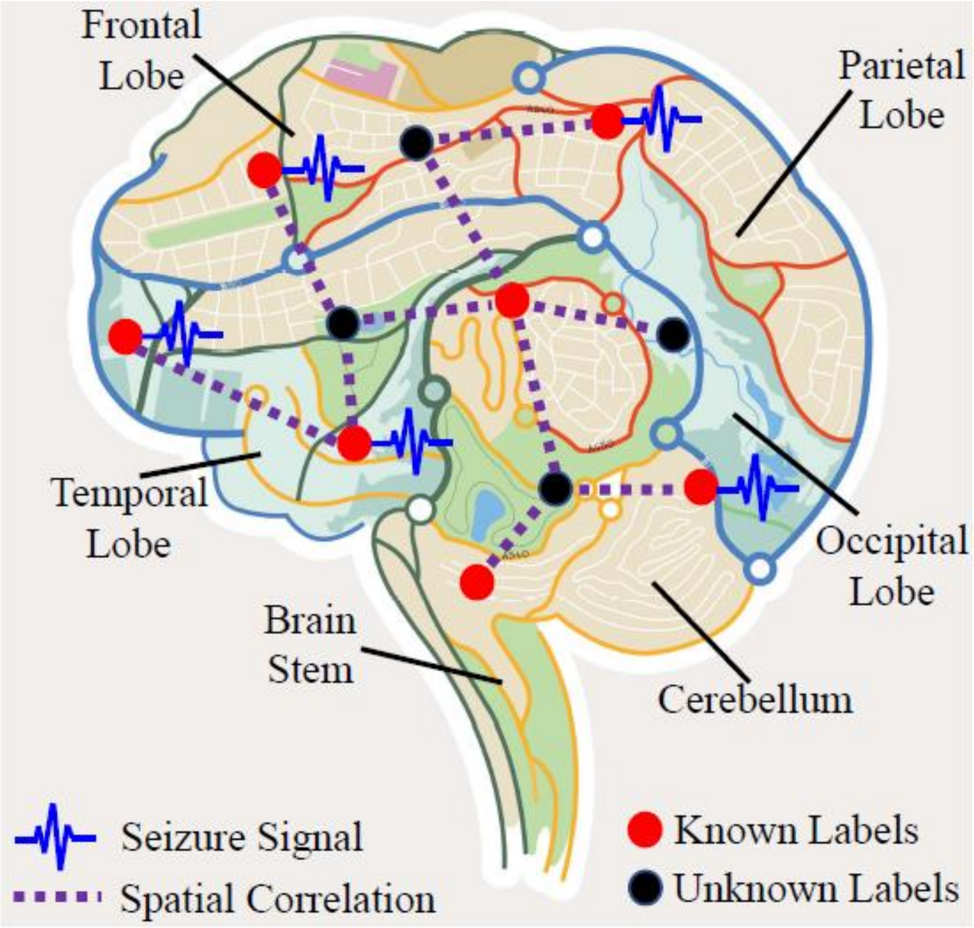
Why Kriging?

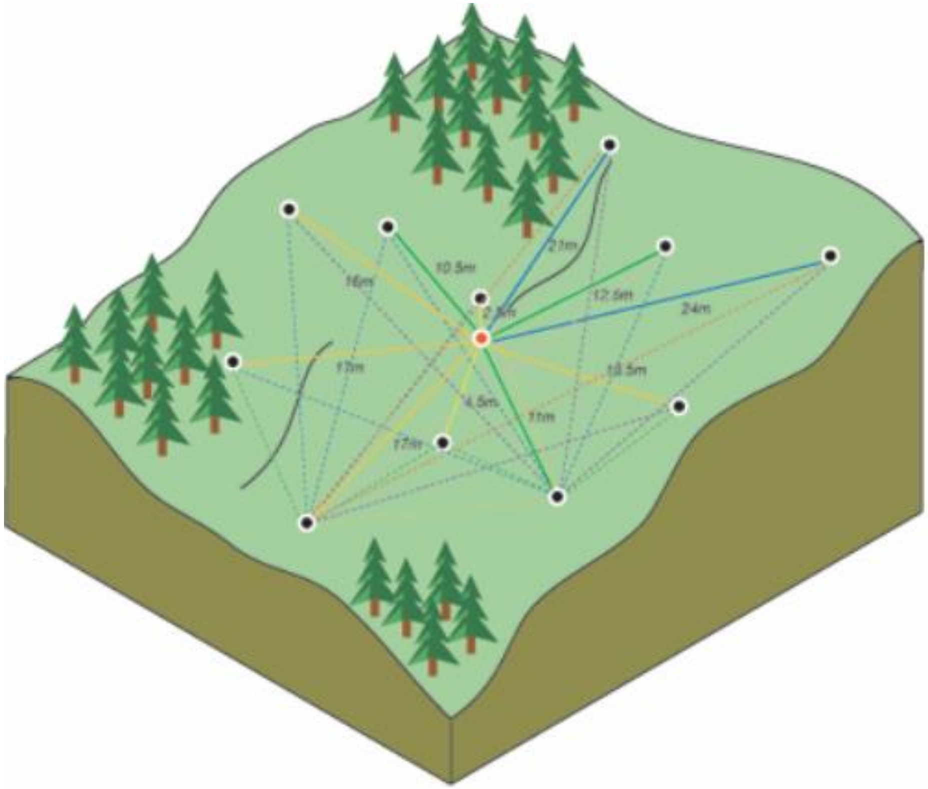
- The modeling of the brain as a geo-spatial map.
- Good performance on small datasets.
- Estimation variance.
- Few hyperparameters.

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





Source:https://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm

Source: I. L. Olokodana, S. P. Mohanty, and E. Kougianos, “Krig-Detect: Exploring Alternative Kriging Methods for Real-Time Seizure Detection from EEG Signal”, in Proceedings of the 6th IEEE World Forum on Internet of Things (WF-IoT), 2020, pp. Accepted.

Kriging Estimates

Given the following set of observations $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, …, $\boldsymbol{x}_n$ as inputs, and $y(\boldsymbol{x}_1)$, $y(\boldsymbol{x}_2)$, …, $y(\boldsymbol{x}_n)$ as outputs, the input-output relationship based on Kriging is given by:

1. $$y(\boldsymbol{x}_i) = \mu + Z(\boldsymbol{x}_i)$$

A residual equation can then be written as:

2. $y(\boldsymbol{x}_0) - \mu_z(\boldsymbol{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\boldsymbol{x}_i) - \mu_z(\boldsymbol{x}_i)$

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

$$y(\boldsymbol{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\boldsymbol{x}_i) + (1 - \sum_{i=1}^{n} \lambda_i)\ \mu_z$$

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

4. $$\boldsymbol{R}^*(\boldsymbol{x}_0) = \sum_{i=1}^{n} \lambda_i \boldsymbol{R}(\boldsymbol{x}_i)$$

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

5. $$\sigma_{est.}^2 = \boldsymbol{E}\{[\boldsymbol{R}^*(\boldsymbol{x}_0) - \boldsymbol{R}(\boldsymbol{x}_0)]^2\}$$

Expanding the equation gives rise to: *6*.

$$\sigma_{est.}^2 = \sum_{i=1}^{n} . \sum_{j=1}^{n} \lambda_i \lambda_j C(\boldsymbol{x}_i, \boldsymbol{x}_j) - 2\sum_{i=1}^{n} \lambda_i C(\boldsymbol{x}_0, \boldsymbol{x}_i) + C(0)$$

The partial derivative of Eqn. *6* . above with respect to λ_i results in:

7. $$\frac{\partial \sigma_{est.}^2}{\partial \lambda_i} = \sum_{j=1}^{n} \lambda_j C(\boldsymbol{x}_i, \boldsymbol{x}_i) - 2. C(\boldsymbol{x}_0, \boldsymbol{x}_i)$$

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

8. $$\sum_{j=1}^{n} \lambda_j C(\boldsymbol{x}_i, \boldsymbol{x}_i) = 2. C(\boldsymbol{x}_0, \boldsymbol{x}_i)$$

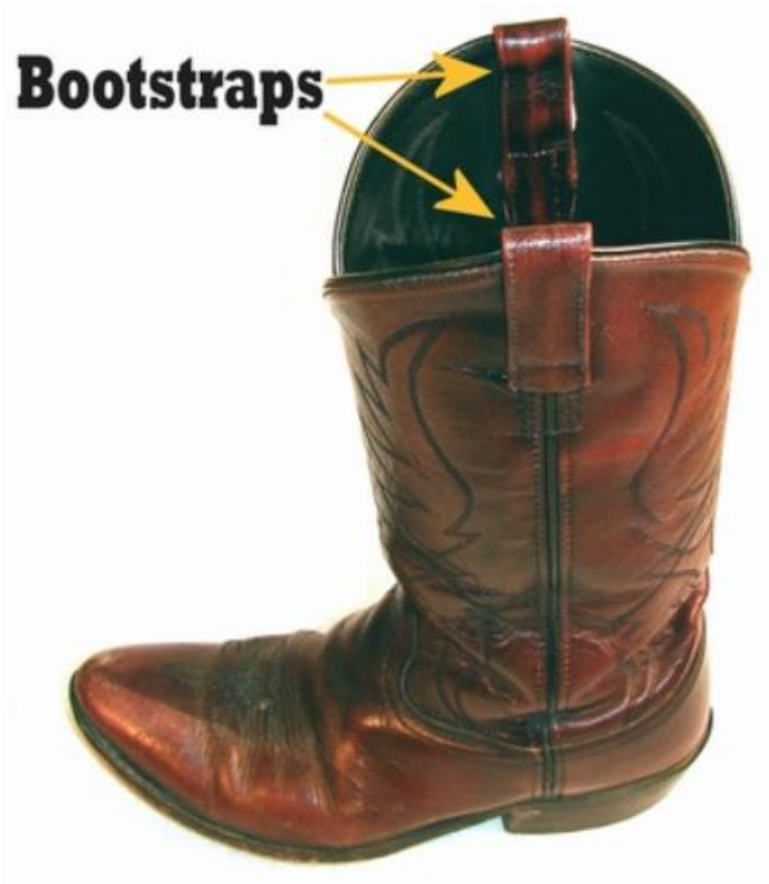
The weights (λ_i) can then finally be obtained by solving Eqn. *8*. Hence, the kriging estimate $y(\boldsymbol{x}_0)$ can be obtained from Eqn *3*.

The covariance function is given by:

9. $$C_x(\boldsymbol{h}) = \sigma_x^2(\boldsymbol{h}) - \gamma(\boldsymbol{h})$$

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

$$\left.\begin{array}{r}\text{draw } x_1, \ldots, x_n \text{ from } P \\ \text{compute } M_1 = g(x_1, \ldots, x_n)\end{array}\right] \quad 1$$

$$\left.\begin{array}{r}\text{draw } x_1, \ldots, x_n \text{ from } P \\ \text{compute } M_2 = g(x_1, \ldots, x_n)\end{array}\right] \quad 2$$

$$\vdots$$

$$\left.\begin{array}{r}\text{draw } x_1, \ldots, x_n \text{ from } P \\ \text{compute } M_B = g(x_1, \ldots, x_n)\end{array}\right] \quad B$$

Get B Statistic

$$M_1$$

$$M_2$$

$$\vdots$$

$$M_B$$

Summarize

Mean

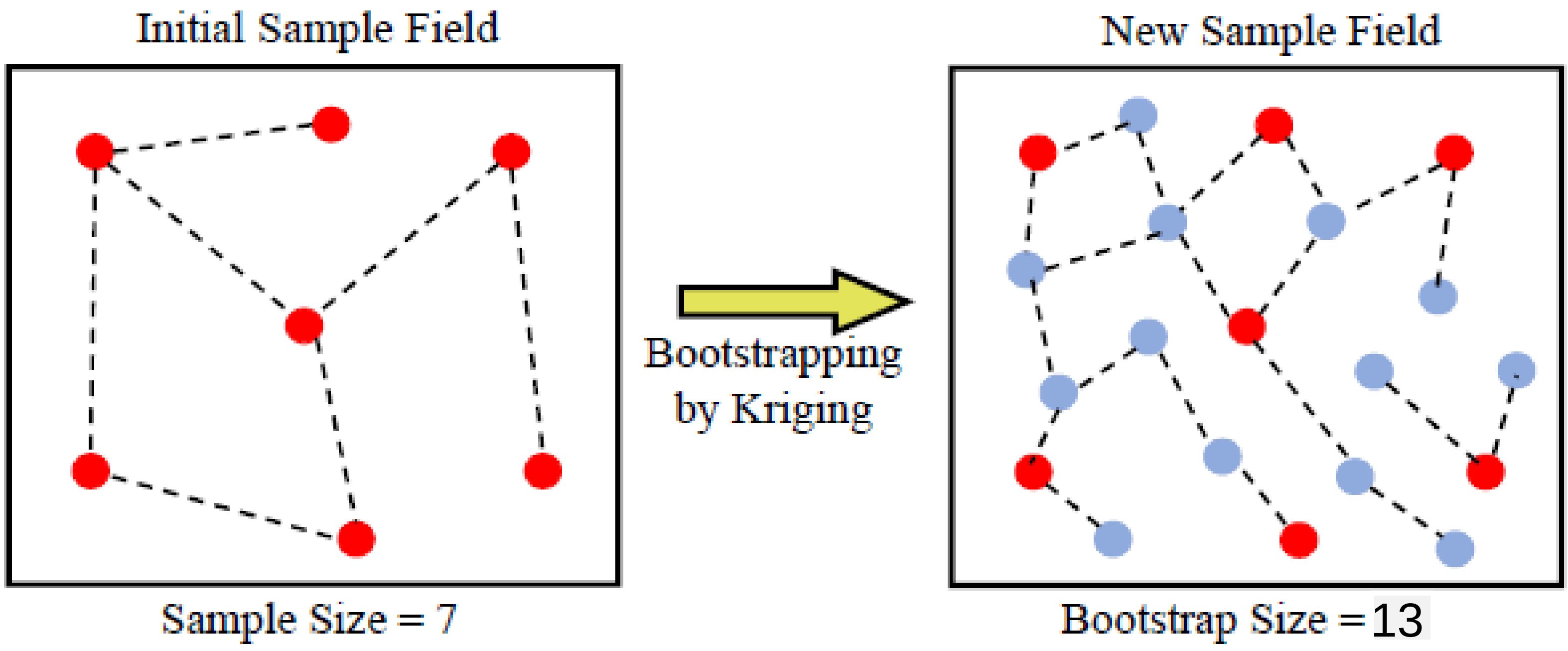
$$\bar{m} = \frac{1}{B} \sum_{j=1}^{B} M_j$$

Variance

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

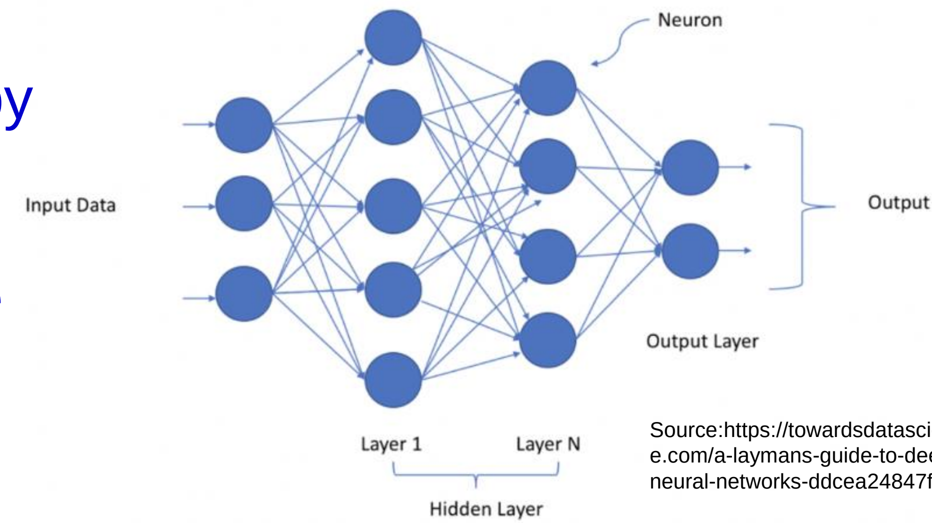
https://towardsdatascience.com/an-introduction-to-the-bootstrap-method-58bcb51b4d60

Bootstrapped Kriging



Deep Neural Network

- The neural network operates by minimizing the cost function.
- The cost function is given by the following expression:



Source:https://towardsdatascience.com/a-laymans-guide-to-deep-neural-networks-ddcea24847fb

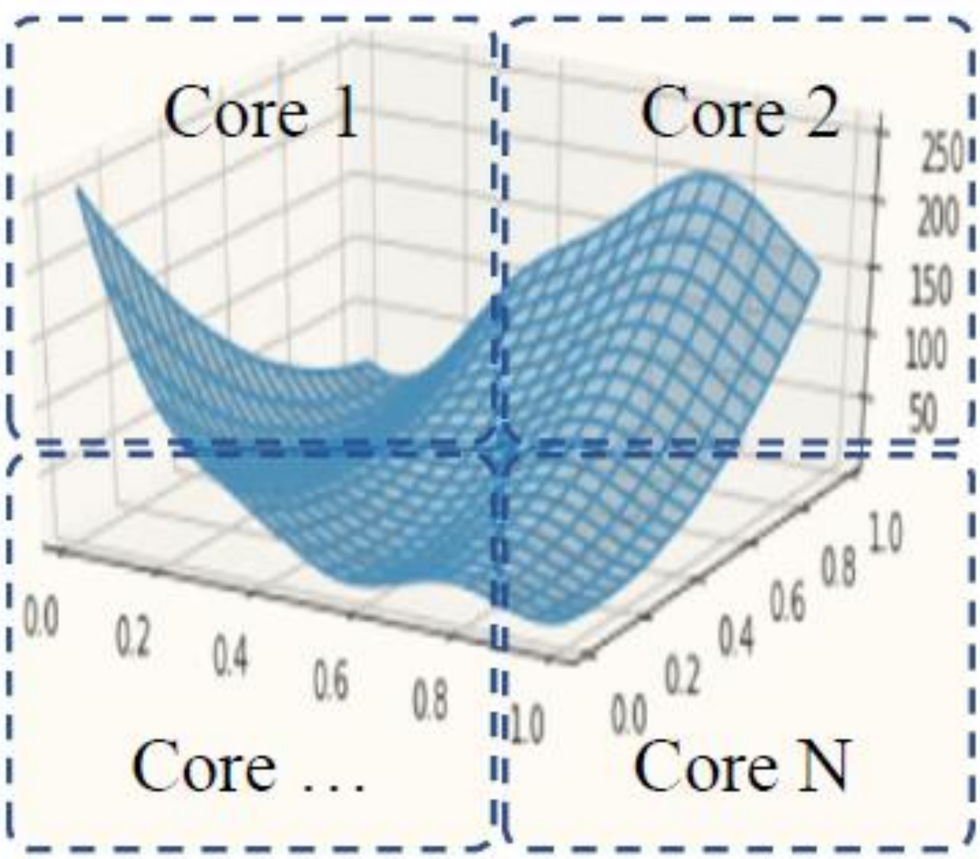
$$J(\omega, b) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)]$$

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

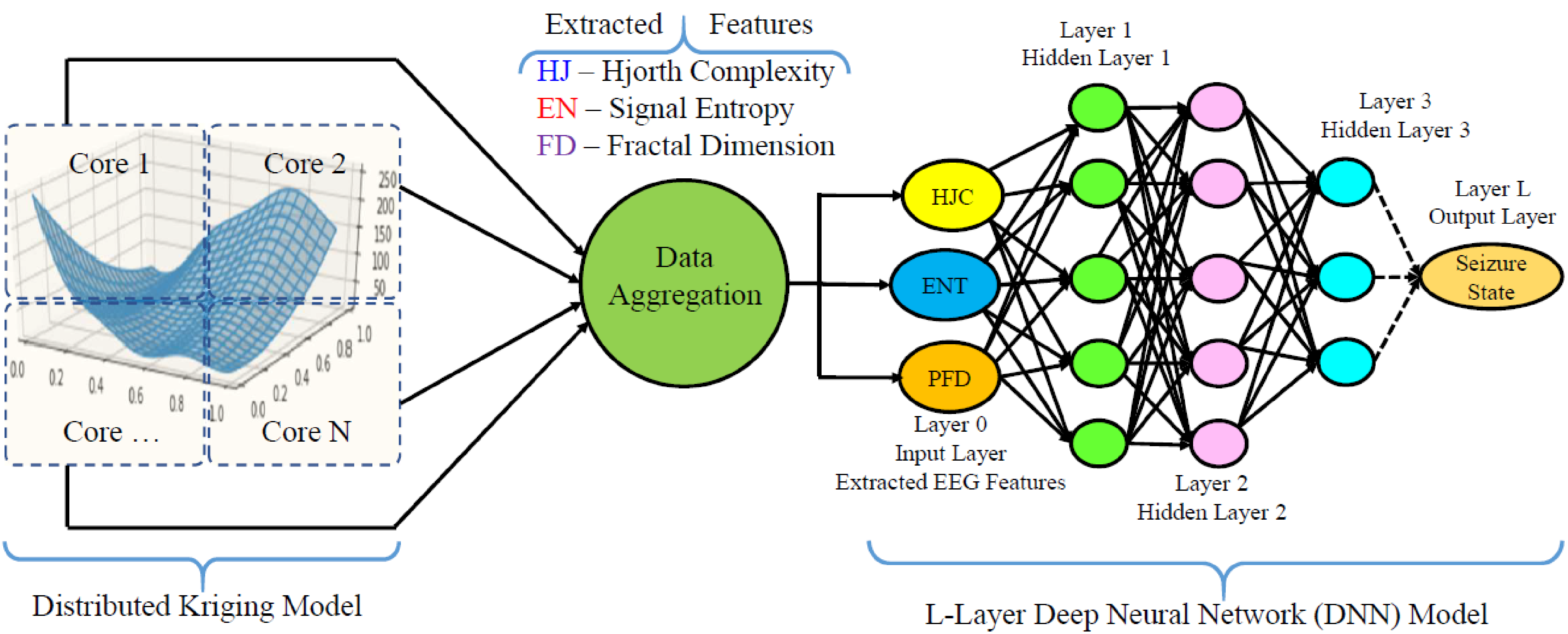
- Where ω refers to weights, b is bias, α is learning rate, n is number of samples. y_i and $\hat{y}_i$ are true are predicted values respectively.

$$\hat{y} = f(\omega^T \boldsymbol{X} + b)$$

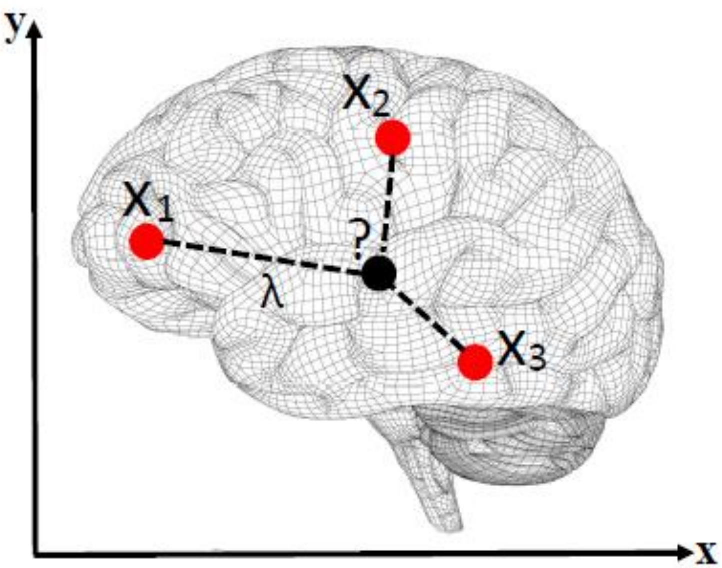
Proposed Distributed Kriging Model



Proposed Distributed Kriging-Bootstrapped DNN Model



Computational Analysis of Distributed Kriging

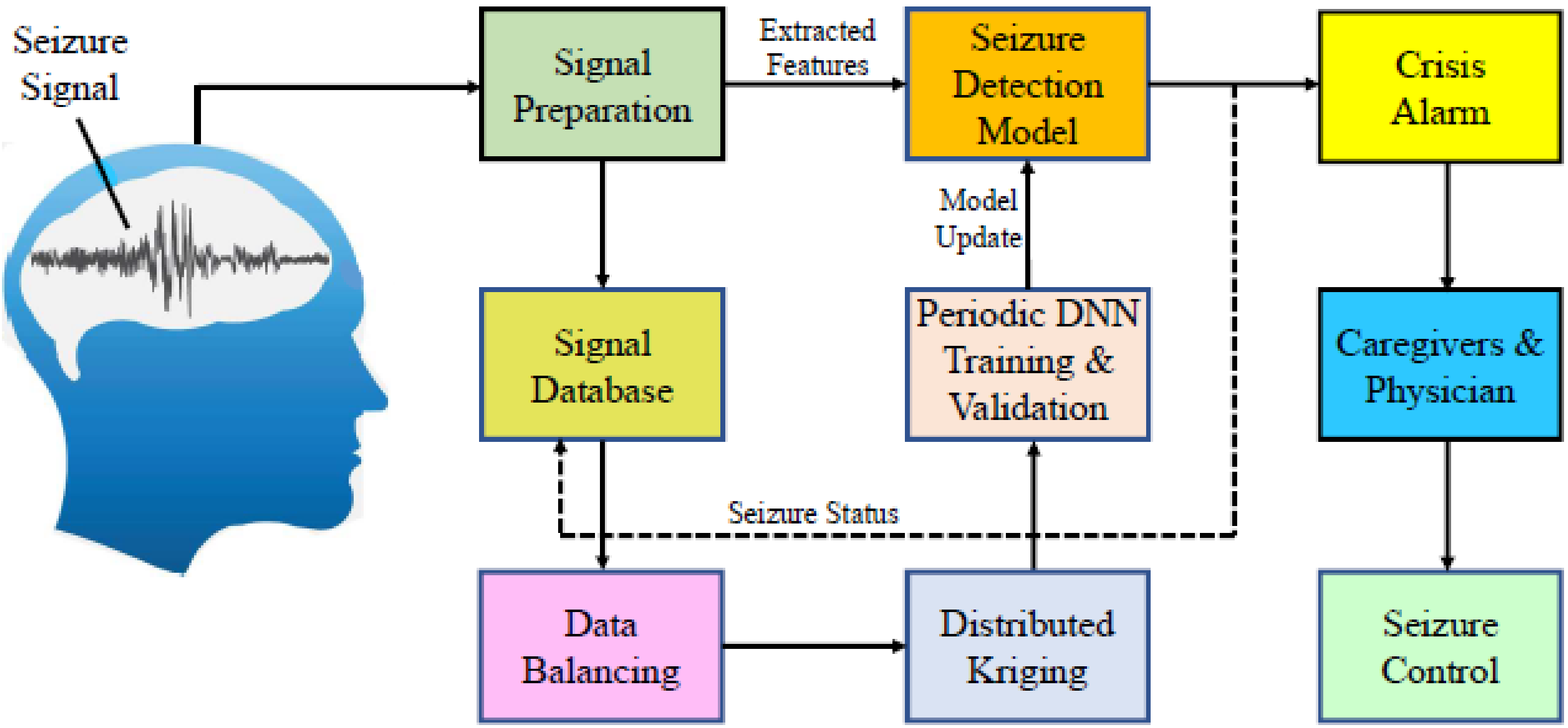


$$\sum_{j=1}^{n} \lambda_j C\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = C(\boldsymbol{x}_0, \boldsymbol{x}_i)$$

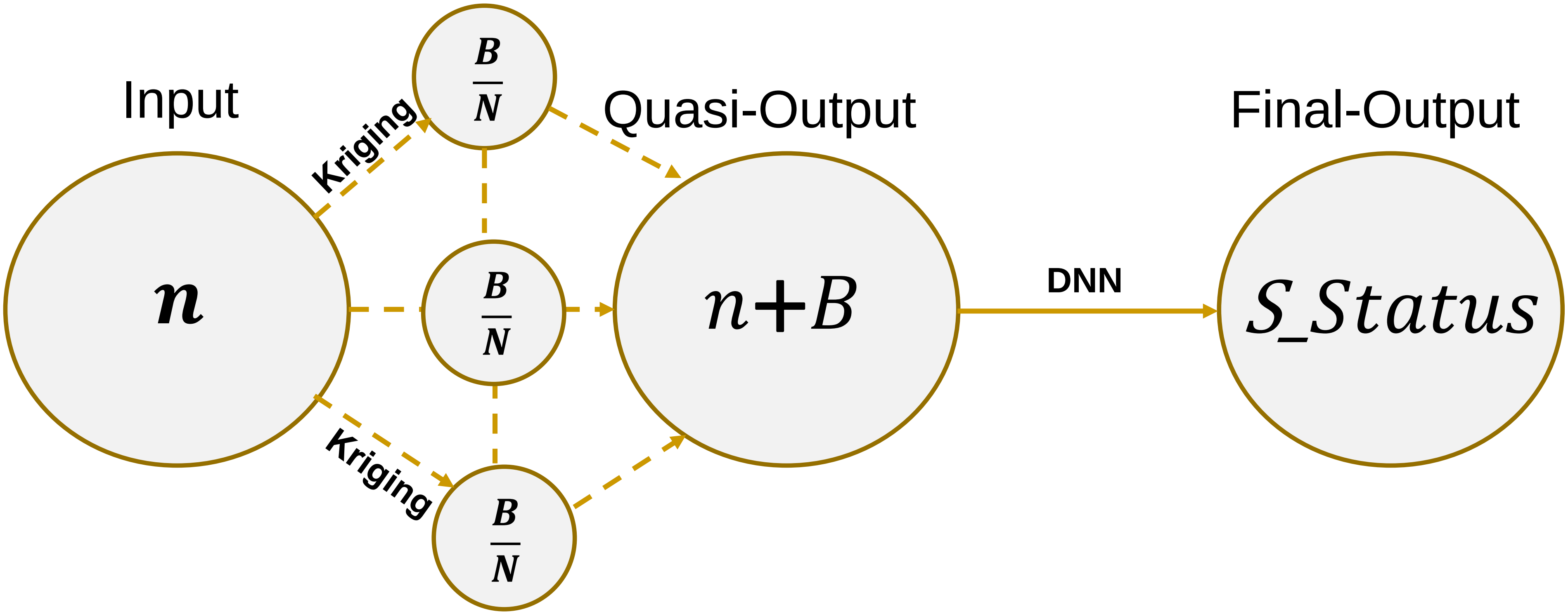
$$\begin{bmatrix} C(\boldsymbol{x}_1,\boldsymbol{x}_1) & C(\boldsymbol{x}_1,\boldsymbol{x}_2) & C(\boldsymbol{x}_1,\boldsymbol{x}_3) \\ C(\boldsymbol{x}_2,\boldsymbol{x}_1) & C(\boldsymbol{x}_2,\boldsymbol{x}_2) & C(\boldsymbol{x}_2,\boldsymbol{x}_3) \\ C(\boldsymbol{x}_3,\boldsymbol{x}_1) & C(\boldsymbol{x}_3,\boldsymbol{x}_2) & C(\boldsymbol{x}_3,\boldsymbol{x}_3) \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \begin{bmatrix} C(\boldsymbol{x}_0,\boldsymbol{x}_1) \\ C(\boldsymbol{x}_0,\boldsymbol{x}_2) \\ C(\boldsymbol{x}_0,\boldsymbol{x}_3) \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} C(\boldsymbol{x}_1,\boldsymbol{x}_1) & C(\boldsymbol{x}_1,\boldsymbol{x}_2) & C(\boldsymbol{x}_1,\boldsymbol{x}_3) \\ C(\boldsymbol{x}_2,\boldsymbol{x}_1) & C(\boldsymbol{x}_2,\boldsymbol{x}_2) & C(\boldsymbol{x}_2,\boldsymbol{x}_3) \\ C(\boldsymbol{x}_3,\boldsymbol{x}_1) & C(\boldsymbol{x}_3,\boldsymbol{x}_2) & C(\boldsymbol{x}_3,\boldsymbol{x}_3) \end{bmatrix}^{-1} \begin{bmatrix} C(\boldsymbol{x}_0,\boldsymbol{x}_1) \\ C(\boldsymbol{x}_0,\boldsymbol{x}_2) \\ C(\boldsymbol{x}_0,\boldsymbol{x}_3) \end{bmatrix}$$

Proposed Fast & Accurate Real-time Seizure Detection Model



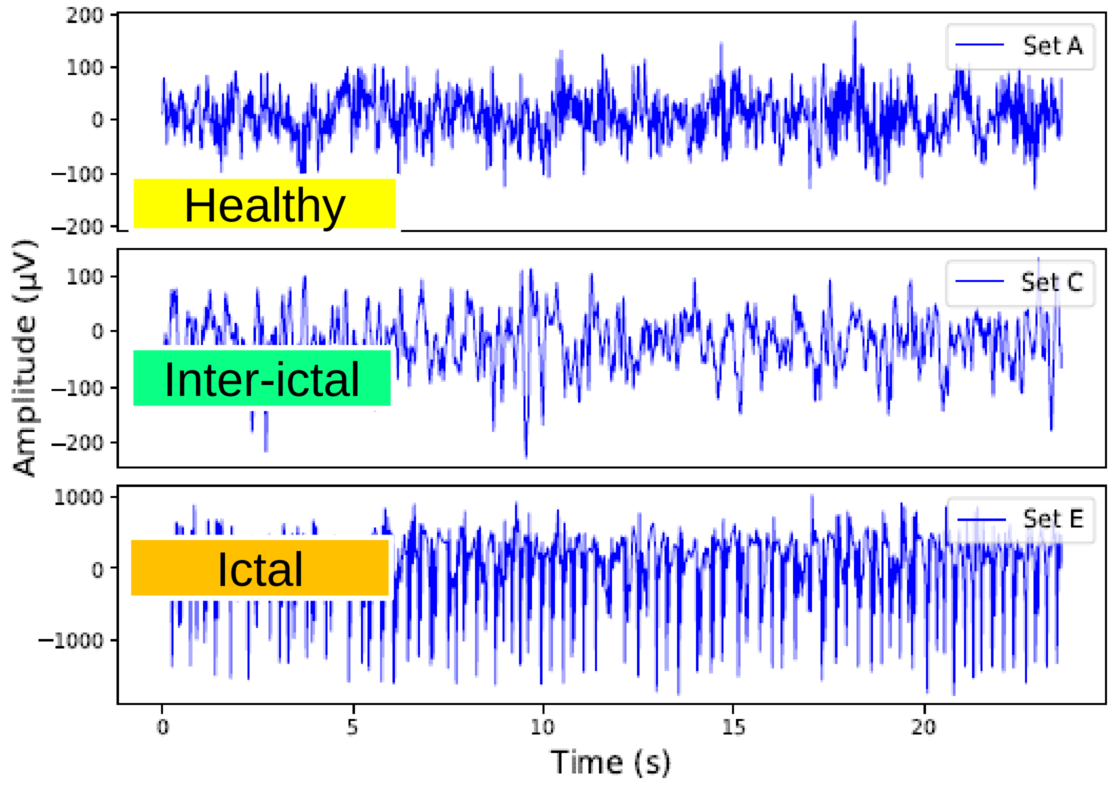
Training or Learning Process



EEG Dataset – Publicly Available

BONN DATASET (DATASET A)

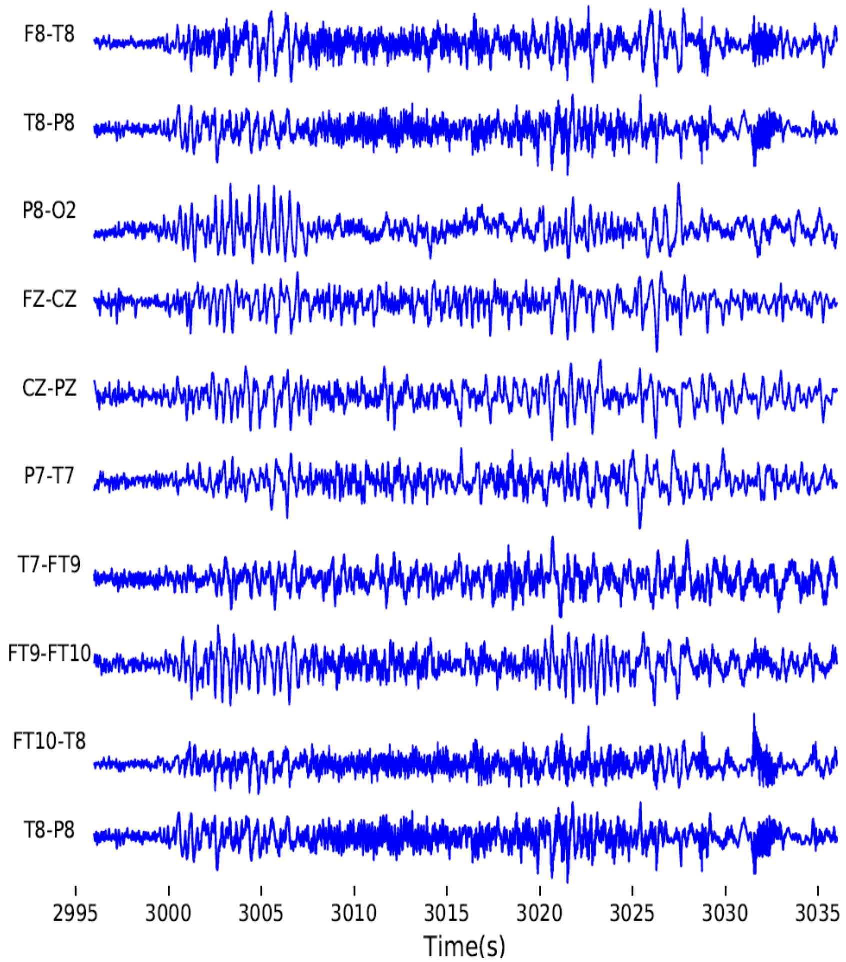
- Consists of 5 sets from A-E. Sets A&B are healthy signals, C&D are inter-ictal signals while E is the only set with ictal signals. Each of the sets comprises 100 EEG segments which were collected with a 128-channel EEG system sampled at 173.61 Hz.



EEG Dataset – Publicly Available

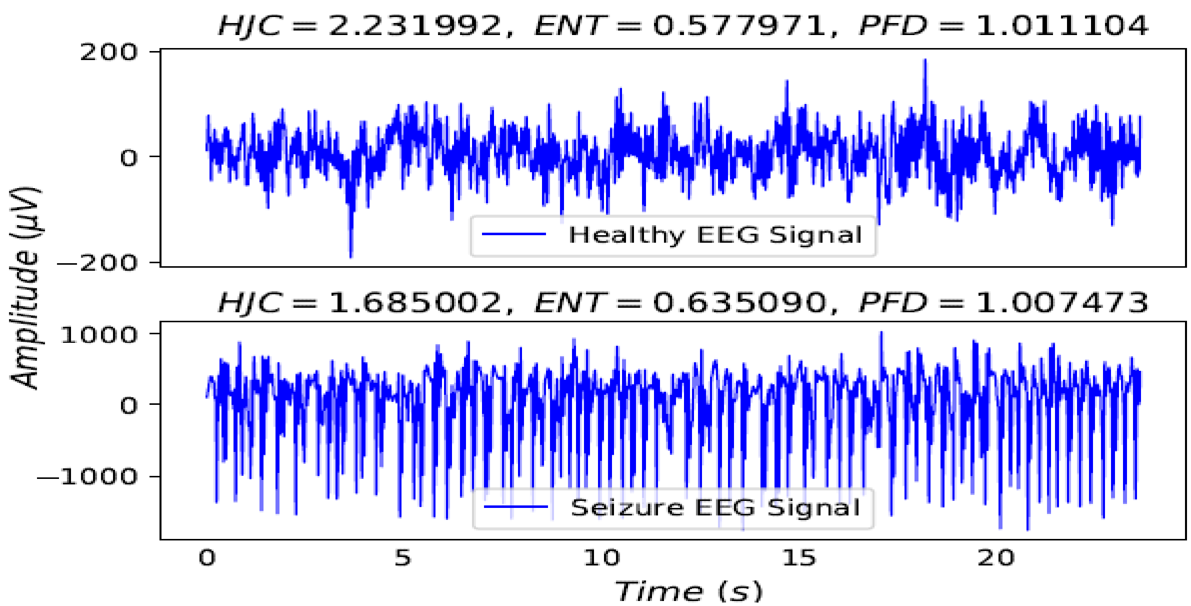
CHB-MIT SCALP EEG DATASET (DATASET B)

- 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.



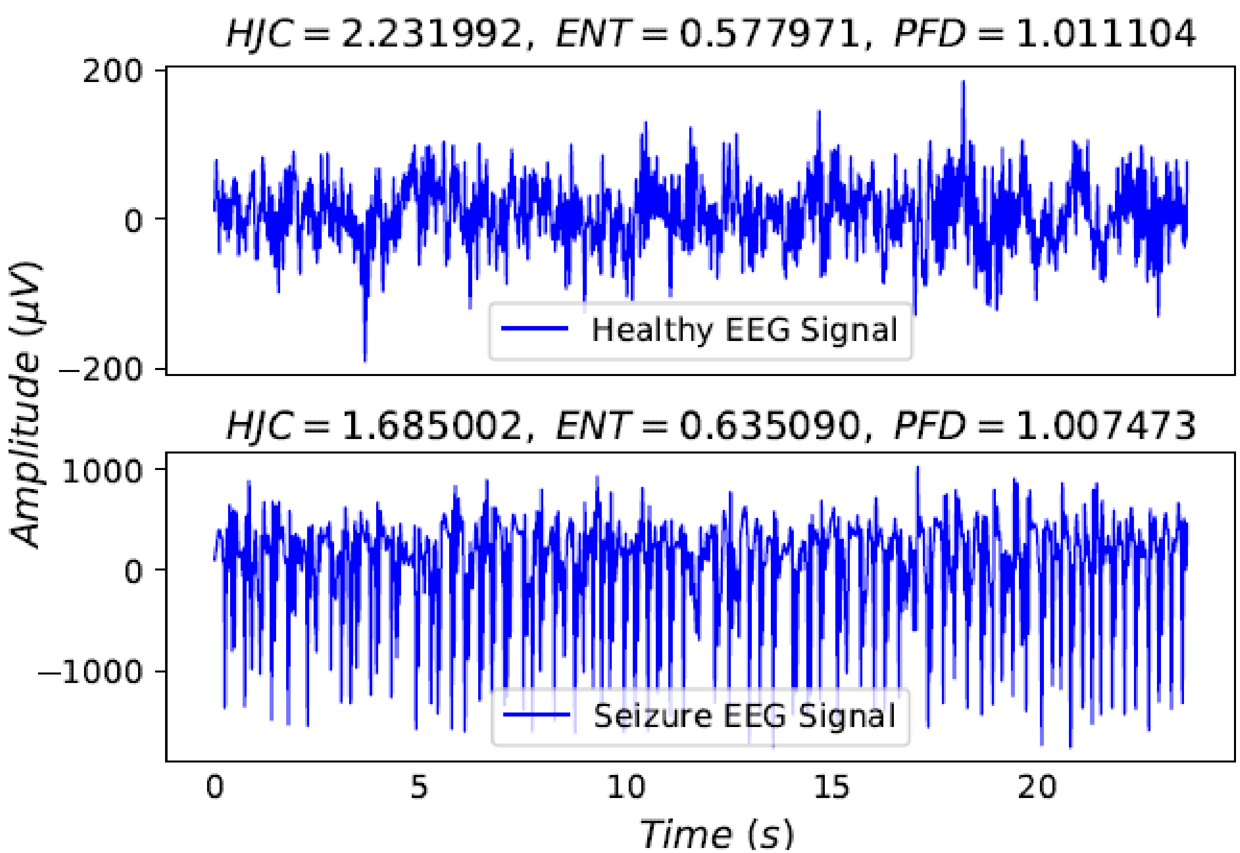
Extracted Features

- 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)



$$HJC = \frac{Mobility(\frac{dy(t)}{dt})}{Mobility(y(t))}$$

where

$$Mobility = \sqrt{\frac{var(\frac{dy(t)}{dt})}{var(y(t))}}$$

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

$$PFD = \frac{\log_e(n)}{\log_e(n)+\log_e(\frac{n}{n+0.4N_\delta})}$$

Experimental Results: Dataset A

- DNN Architectural Specification.

DNN Specifications	Values
No of Layers	4
Hidden Units	5, 5, 3
Hidden L. Activation	Rectified Linear Unit (ReLU)
Output L. Activation	Sigmoid Function
Initialization Method	Xavier Initialization
Optimization Method	Adaptive Momentum

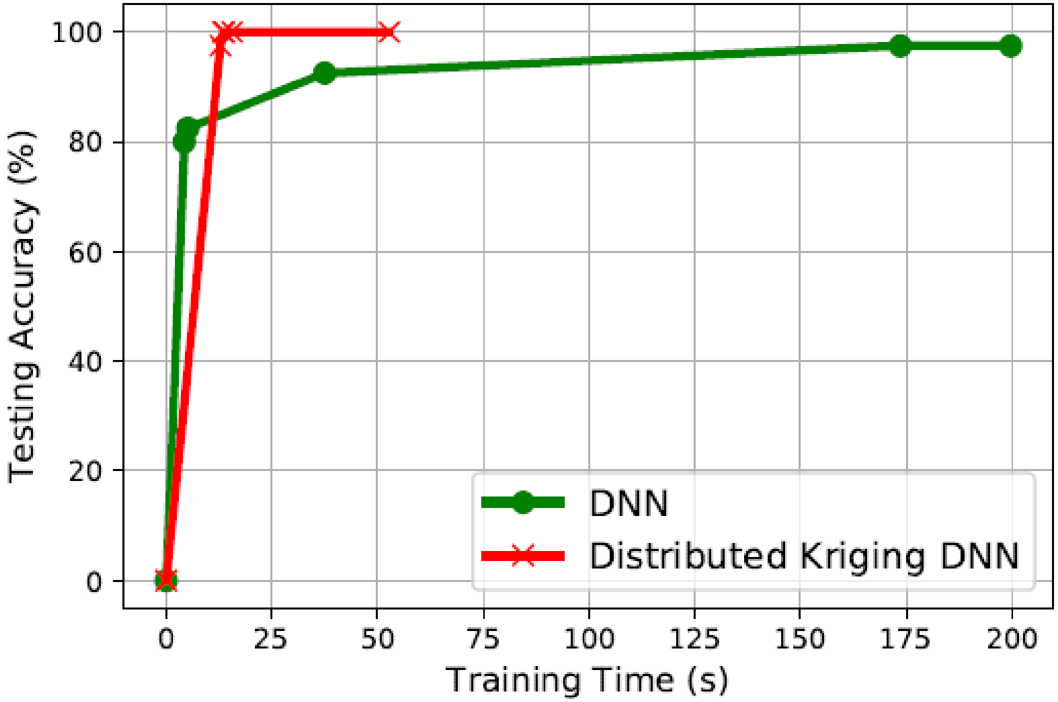
- Distributed kriging-bootstrapped DNN model performance with 10,000 samples using Dataset A.

Count	Training Accuracy	Testing Accuracy	Training Epochs	Training Time
1	99.14%	97.50%	500	12.80s
2	99.76%	100.00%	800	13.46s
3	99.84%	100.00%	1000	13.75s
4	99.92%	100.00%	1500	15.56s
5	99.92%	100.00%	10000	52.72s

Experimental Results: Dataset A

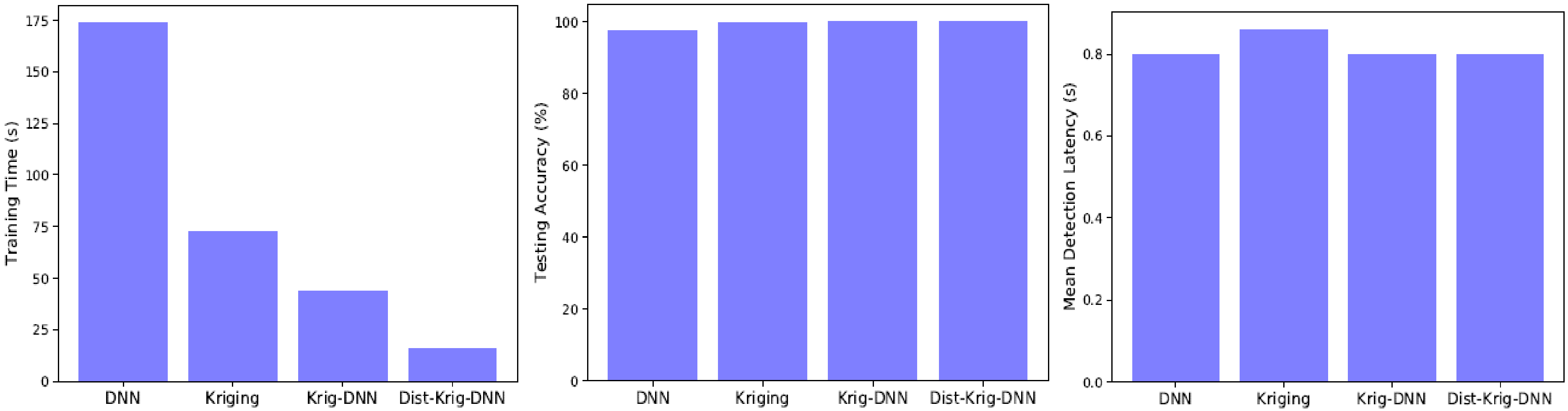
- Training Time reduced by 91%

Models	DNN	Ordinary Kriging	Kriging DNN	Distributed Kriging DNN
Tr. Data Size	10000	2000	10000	10000
Tr. Epochs	45000	NA	1500	1500
Learning Rate	0.00001	NA	0.001	0.001
Training Acc.	99.99%	100.00%	99.92%	99.92%
Testing Acc.	97.50%	99.78%	100.00%	100.00%
Training Time	173.57s	72.24s	43.83s	**15.56s**



Experimental Results: Dataset A

Models	Detection Latency
DNN	0.80s
Ordinary Kriging	0.86s
Krig-DNN	0.80s
Dist-Krig-DNN	0.80s



Experimental Results: Dataset B

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

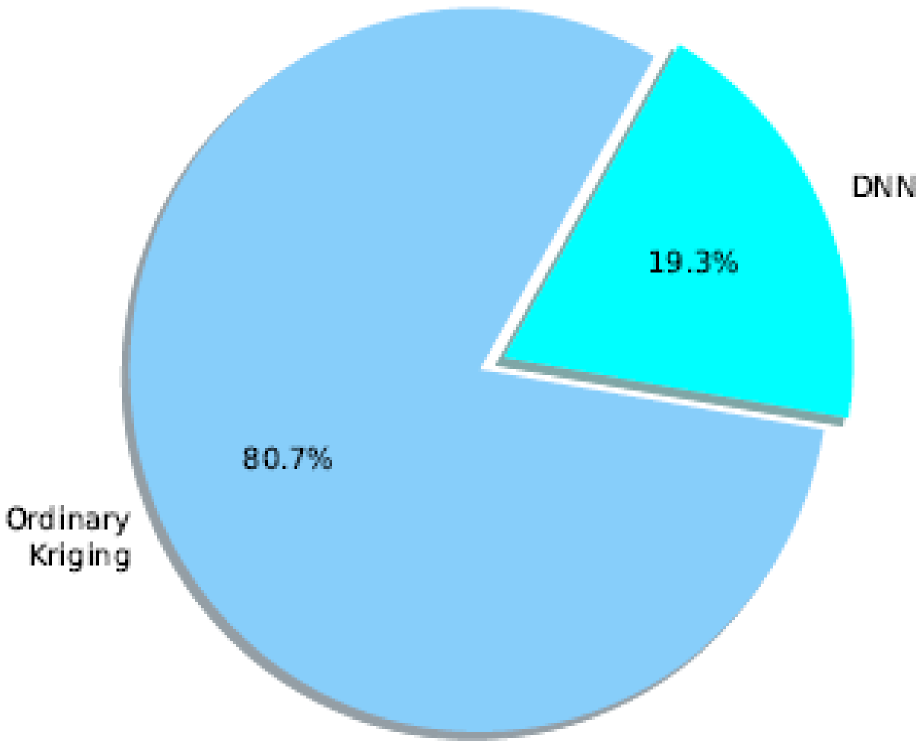
Models	Channel Type	No of Channels	Training Accuracy	Testing Accuracy
Kriging	Single	1	68.00%	59.00%
Kriging	Multiple	23	99.70%	89.00%
Dist-Krig-DNN	Single	1	100.00%	98.53%

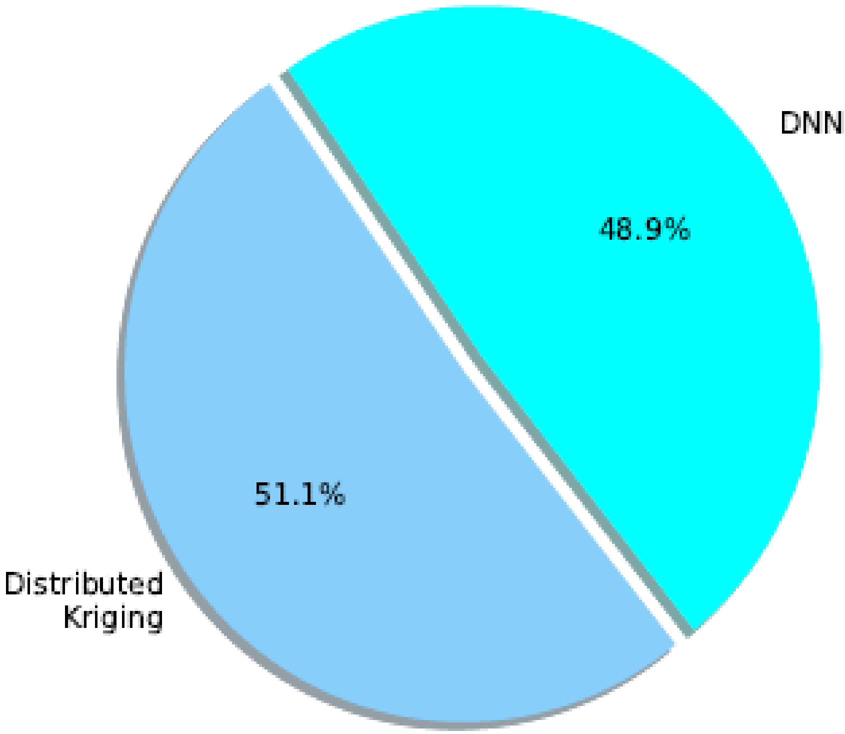
- Comparing best performances for basic DNN and distributed Kriging-bootstrapped DNN models using Dataset B.

Models	Training Accuracy	Testing Accuracy	Training Epochs	Training Time
DNN	99.97%	98.53%	10000	42.03s
Dist-Krig-DNN	100.00%	98.53%	500	7.05s

- 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





Comparison with Related Works

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

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

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

1. I. L. Olokodana, S. P. Mohanty, and Elias Kougianos, "Ordinary-Kriging Based Real-Time Seizure Detection in an Edge Computing Paradigm", *in Proceedings of the 38th IEEE International Conference on Consumer Electronics (ICCE),* 2020, DOI: 10.1109/ICCE46568.2020.9043004.

2. I. L. Olokodana, S. P. Mohanty, and Elias Kougianos, "Krig-Detect: Exploring Alternative Kriging Methods for Real-Time Seizure Detection from EEG Signal", *in Proceedings of the 6th IEEE World Forum on Internet of Things (WF-IoT),* 2020, pp. Accepted.

3. I. L. Olokodana, S. P. Mohanty, and Elias Kougianos, "Kriging-Bootstrapped DNN Hierarchical Model for Fast, Accurate Seizure Detection from EEG Signals", *in Proceedings of the 6th IEEE World Forum on Internet of Things (WF-IoT),* 2020, pp. Accepted.

4. I. L. Olokodana, S. P. Mohanty, and Elias Kougianos, "Distributed Kriging-Bootstrapped DNN Model for Fast, Accurate Seizure Detection from EEG Signals", *in Proceedings of the 19th IEEE-CS Symposium on VLSI (ISVLSI),* 2020, pp. Accepted.

THANK YOU