Long Short-Term Memory (LSTM) based Epileptic Seizure Recognition

Sunil Kumar D.S., PhD
Department of Computer Science
Mangalore University, Mangalore, India

Mamatha Mallesh
Assistant Professor, Department of ECE
Jnanavikas Institute of Technology, Bidadi, India

Bharath K.N.
Assistant Professor, Department of ECE,
DSATM, Bangalore, India

Susmitha B.C.
Assistant Professor, Department of ECE,
Jnanavikas Institute of Technology, Bidadi, India

Kiran
Assistant Professor, Department of ECE,
Vidyavardhaka College of Engineering, Mysuru, India

ABSTRACT
Epilepsy is the second most common brain disorder after migraine; automatic detection of epileptic seizures can considerably improve the patients' quality of life. Current Electroencephalogram (EEG)-based seizure detection systems encounter many challenges in real-life situations; EEG data are prone to numerous noise types that negatively affect the detection accuracy of epileptic seizures. To address this challenge, we propose a deep learning-based approach that learns the discriminative EEG featuresof epileptic seizures and to distinguish between the different types of patient recordings. More specifically, we aim to tackle this issue by using a Long Short-Term Memory network, and explore the capabilities of this model.

General Terms
Recognition, Seizure detection

Keywords
Epileptic Seizure, Recognition, LSTM, EEG data

1. INTRODUCTION
EPILEPSY is a chronic neurological disorder of the brain that affects people of all ages. Approximately 70 million people worldwide have epilepsy, making it the second most common neurological diseases after migraine [1]. The defining characteristic of epilepsy is recurrent seizures that strike without warning. Symptoms may range from brief suspension of awareness to violent convulsions and sometimes loss of consciousness [2]. Epileptic seizure detection plays a key role in improving the quality of life of epileptic patients. Electroencephalogram (EEG) is the prime signal that has been widely used for the diagnosis of epilepsy. The visual inspection of EEG is unfortunately labor and time-consuming. Also, around 75% of people with epilepsy live in low- and middle income countries and cannot afford consulting neurologists or practitioners [3]. Those limitations have encouraged scholars to develop automatic EEG-based seizure detection systems. A vast number of methods have been developed for automatic seizure detection using EEG signals. Extracting features that best describe the behavior of EEGs is of great importance for automatic seizure detection systems' performance. Several feature extraction and selection techniques have been reported in the literature. Most of them use hand-wrought features in the time-domain [4], [5], frequency-domain [6]–[8], time frequency domain [9]–[12] or sometimes in a combination of two domains [13]. However, these domain-based methods encounter three main challenges. First, they are sensitive (not robust enough) to acute variations in seizure patterns. This is because the EEG data is non-stationary and its statistical features change across different subjects and over time for the same subject. Secondly, EEG data acquisition systems are very susceptible to a diverse range of artifacts such as muscle activities, eye-blinks, and environmental white noise. All these sources of noise can alter the genuine EEG features and hence seriously affect the performance accuracy of seizure detection systems. The authors of [14] have studied the impact of high noise levels on the recognition performance of epileptic seizures. It is worth highlighting that detecting seizures from noisy EEG data corrupted with a medium-level noise has resulted in a drop of 10% in the seizure detection accuracy [14]. Finally, most existing seizure detection systems have been trained on small-scale EEG datasets collected from few specific patients, making them less practical in clinical applications. Due to the nonstationary nature of EEG, especially during seizures, entropy measures have attracted more attention in the field [15-17].

The data, published on Bonn University's Epileptology department website, presents Electroencephalogram (EEG) recording of 500 individuals. For each individual, brain activity was recorded for a duration of 23.5 seconds; these recordings are represented by 4096 evenly-spaced, consecutive data points (i.e every 0.0057 seconds). Each row of the dataset, representing an individual's recording, also has a column with the classification of the recording. The five labeled datasets (A, B, C, D, E) are presented below along with their corresponding target classes:

Our objective for this method is to correctly predict these five target classes using an LSTM Neural Network. In order to highlight the capabilities of an LSTM layer, three classification problems will be tackled where the LSTM's performance will be compared to that of a regular deep neural network:

- A two-class problem that classifies whether a patient is having a seizure or not at the time of recording.
- A three-class problem that requires the classifier to distinguish between a patient that is having a seizure, a

patient that is between seizures (*inter-ictal*), and a healthy patient.

- A five-class problem that aims to classify all five classes, meaning it should be able to distinguish between a patient that is having a seizure, a patient that is between seizures, and a healthy patient. Additionally, it will be able to determine in which part of the brain the recording is made (Sets C & D) and whether the patient's eyes are open or closed (Sets A & B).

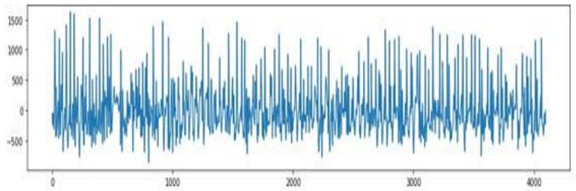

Figure 1: Set A - Class 4: EEG recording of a non-epileptic awake patient with eyes open

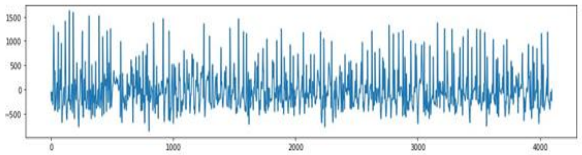

Figure 2: Set B - Class 3: EEG recording of a non-epileptic awake patient with eyes closed.

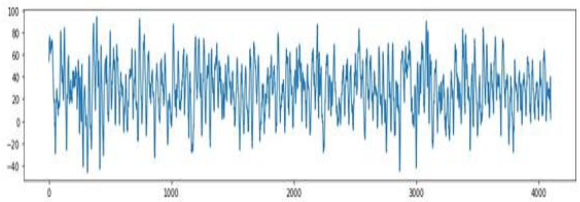

Figure 3: Set C - Class 2: EEG recording of an epileptic patient during seizure free period using electrodes implanted in the brain epileptogenic zone.

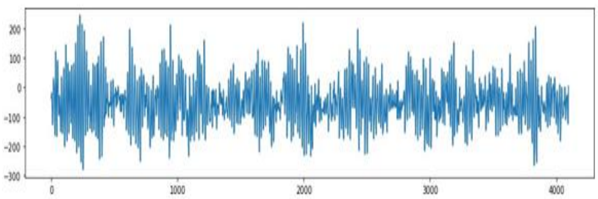

Figure 4: Set D - Class 1: EEG recording of an epileptic patient during seizure free period from the hippocampal formation of the opposite hemisphere of the brain from C.

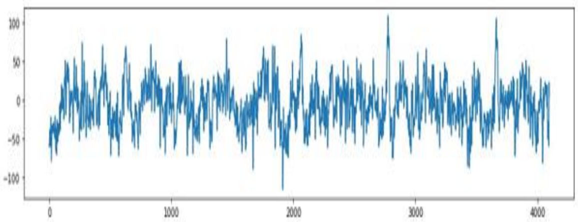

Figure 5: Set E - Class 0: EEG recording of a patient experiencing an active epileptic stroke.

2. EVALUATION MEASURE

To measure the degree of success of our classifiers we have to look at the three problems separately. Especially in medical cases correct classification is of high importance (compared to marketing problems for instance), due to the direct impact on people.

For the two-class classification problem we have a very unbalanced dataset at hand. A high accuracy could still mean, that we classify those with an active seizure incorrectly, without representing this impact adequate when optimizing for accuracy. Therefore, the correct measure would be precision or recall. As it is an ethical question though, whether it is more important to classify someone having a seizure correctly, or someone not having one incorrectly, we opted to include the confusion matrix that takes both into consideration. But we display the accuracy in the other graphs to keep the results comparable. For the three-class classification problem the data is almost balanced in the sense that wehave 300 patients with, and 200 without epilepsy. We consider this an edge case, where there are equalarguments for and against accuracy, but opted for easier interpretability for accuracy. For the five-class classification problem, we have a perfectly balanced dataset, therefore, accuracy is in our perspectivethe best measure. To reduce the impact of randomness, all results are averaged over five seeds.

2.1 Data Preparation

As the data is considered clean, and the entire dataset has to be used for the question at hand, no modifications were applied to the data. As the dataset is balanced with 100 cases for each category over- or under-sampling techniques were not applied. Enriching the dataset to reduce overfitting in thiscase is from our perspective not possible, as the patterns in the EEG are of high importance to spot epileptic behavior. Trying to replicate these patterns from the existing samples would be prone to errors as the differences between the samples from the five categories are very hard to spot.

2.2 Two-Class Classification

The most basic classifier for this dataset would be able to distinguish between an individual experiencing a seizure and one who is not. For this model, the target classes have been reduced to patients that are currently experiencing a seizure (label 1: Set E), and patients that arenot experiencing a seizure at the time of recording (label 0: Sets A, B, C, D). Our model (*LSTM*) is tested against a regular deep neural network model (*NN*) so as to compare the performance of both models on this simplified task. Figures 6 and 7 show the validation accuracies and losses of our LSTM model and the NN model (averaged over 5 seeds). We notice that although the LSTM outperforms the NN model, the results are fairly close. We canconclude from this that on a simplified binary problem, the LSTM doesn't deliver significantly more performance than a regular neural network.

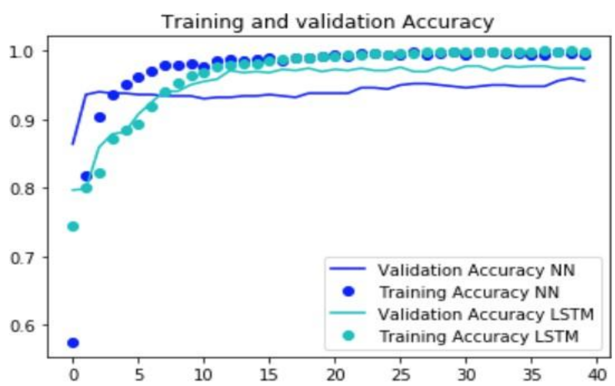

Figure 6 - Training and validation accuracy

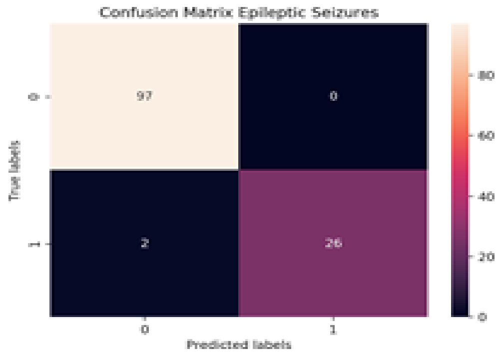

Figure 7 - confusion matrix two class classifications

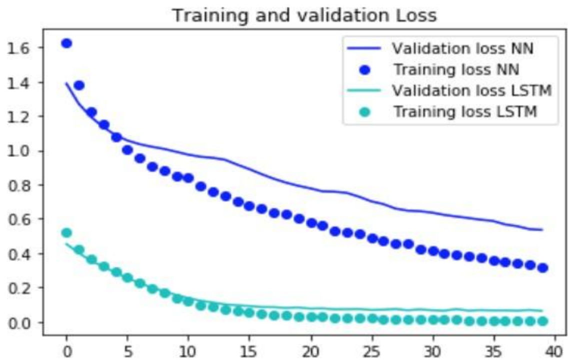

Figure 8 - training and validation loss

2.3 Three-Class Classification

In the three class-problem it becomes clearer that sharing time related information between neurons significantly increases performance and give the LSTM an advantage. Not only does the LSTM perform better, it has much better generalization ability than the regular neural network. We can conclude that the LSTM is appropriate to use for more complex time-series classifications. Figure 9 shows us the confusion matrix of our LSTM; we notice that our model can correctly distinguish between non-seizure patients (label 0), inter-ictal patients (label 1), and seizure patients (label 2).

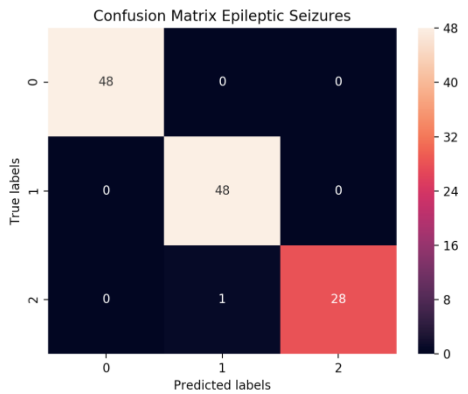

Figure 9 - Confusion matrix three class classification

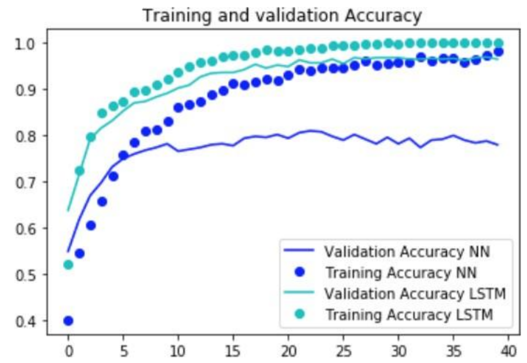

Figure 10 - training and validation fitness

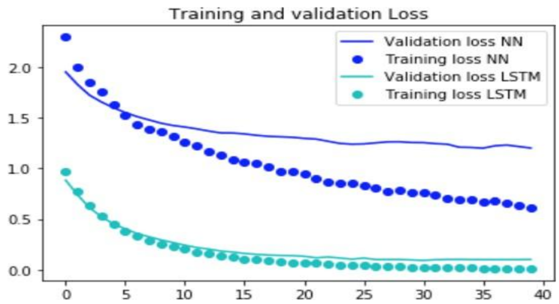

Figure 11 - training and validation loss

The LSTM performs better than the normal neural network on the five-classproblem; this was expected given the results of the three-class problem.However, we noticed that while the LSTM still performs well, there is a significant drop in generalization ability (i.e the LSTM overfits). This holdseven more for the generalization ability of the NN. In this section we will further explore the LSTM through a series of optimization techniques, architectural changes, and feature engineering techniques will be implemented and tested in order to further increase its performance and reduce overfitting.

2.4 Five Class- Classification

We notice from the confusion matrix in figure 14 that the model is mainly having trouble distinguishing between classes *1* & *2*, these correspond to the '*recording of an epileptic patient during seizure free period using electrodes implanted in the brain epileptogenic zone*' and the '*recording of an epileptic patient during seizure free period from the hippocampal formation of the opposite hemisphere of the brain from C*'. In other words, while our model can correctly classify inter-ictal patients, it is having difficulties distinguishing between the different parts of the brain that are being recorded.

It seems our model faces two main issues; overfitting and the inability to distinguish between the two types of inter-ictal recordings. In the following sections of this report, we will further explore the architecture and parameters of our model in order to try and improve it.

3. METHODOLOGY

In the previous part of the report we used a working configuration of the NN and LSTM model to compare them on each respective problem. We did not fine-tune the respective model for this part of the report. In the following part we outline elements of the LSTM that we further addressed when fine-tuning the model. The LSTM consisted of a first LSTM layer with 100 neurons, followed by a 10% dropout, followed by a Time Distributed Dense layer with 50 neurons and a Global average pooling layer. The last layer has the number of neurons fitting to the problem. Further details are in the code.

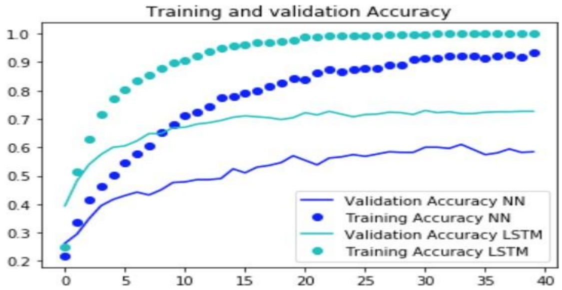

Figure 12 - training and validation fitness

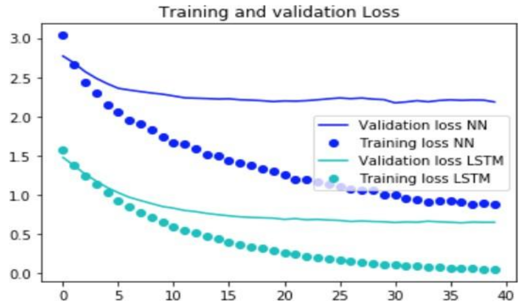

Figure 13 - training and validation loss

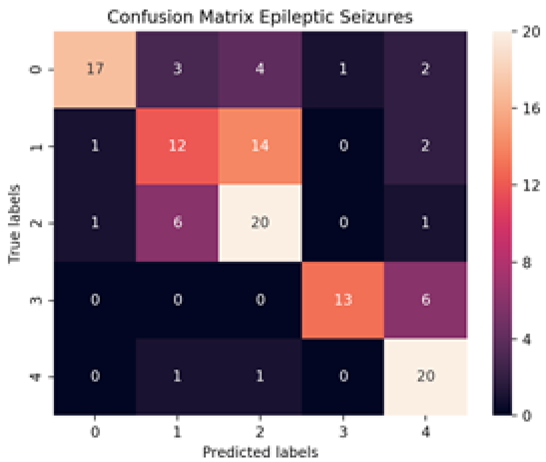

Figure 14 - confusion matrix five class classifications

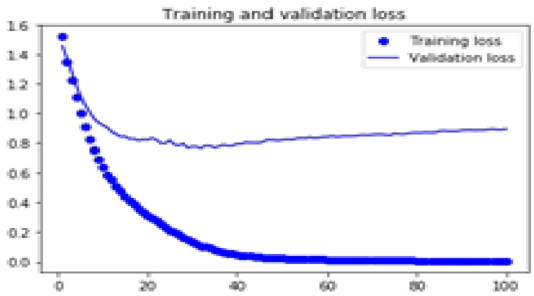

Figure 15- training and validation loss

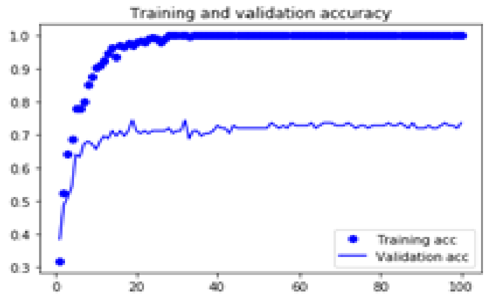

Figure 16 - training and validation fitness

3.1 Epochs

Running our basic LSTM implementation over 100 epochs highlights that the appropriate number of epochs used for our system is at around 40 epochs. In Figure 15, we see our validation accuracy stabilizing at around 40 epochs whilst on the right-hand side we observe that beyond 40 epochs, an increasing dispersion between training loss and validation loss arises, suggesting we would be increasingly overfitting.

3.2 Pooling

Global Pooling layers in convolutional networks play an extremely important role as they reduce thedimension of data from the prior layer through the combination of its neuron clusters into a single neuron in the following layer.

3.3 Global Max Pooling

Global Max pooling on the other hand extracts the maximum value from a cluster of neurons in the previous layer. Perfect testaccuracy is achieved at around 20 epochs and we see our validation accuracy stabilizing at around 55% at 33 epochs and beyond. We observe a final training/validation loss spread of roughly 1.1

3.4 Global Average Pooling

Applying the Global Average pooling layer extracts the average value from a cluster of neurons in the previous layer to then in turn create the appropriate output. Our validation accuracy terminates at just above 70% whereas we observe convergenceto a perfect test accuracy at around 25 epochs. A final training vs validation loss spread of around 0.8 still provides some evidenceof overfitting.

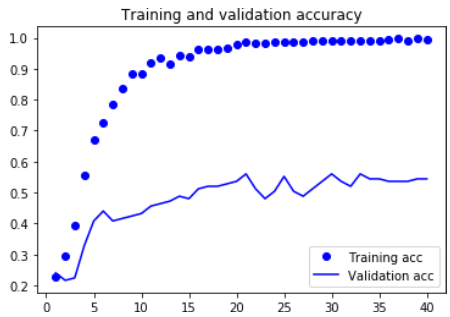

Figure 17 - training and validation fitness

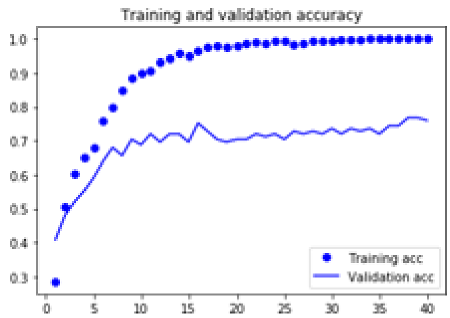

Figure 18 - training and validation fitness

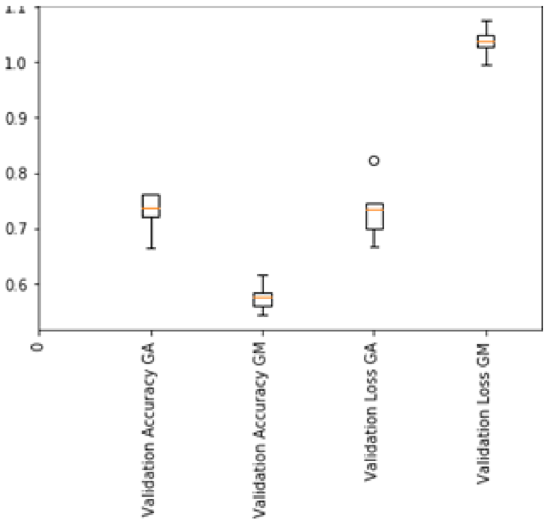

Figure 19: Results comparison polling

3.5 Time steps & Dimension Sizes

Each row of the epileptic seizure dataset represents a 23.5sec EEG recording of an individual; this data is represented as 4096 evenly-spaced, consecutive points. Although it is possible to create an LSTM neural network with 4096 input variables, our belief is that this data can be aggregated into a series of shorter time dimensions; aggregating the time information according to smaller *timesteps* would reduce computational time and create a simpler model. Our parameter *timesteps* represents the number of times the row data will be divided (i.e the number of input variables our LSTM model will receive). For example, if we choose 128 timesteps, our data will be split into 128 input variables, each containing 32 data points (4096/128 = 32).

Our initial model, the one that was tested against a normal neural network for each classification problem, was divided into 64 timesteps; this means the LSTM model received as inputs time blocks corresponding to 64 consecutive recordings. This division was arbitrary; in reality, our time series can be divided into can be divided by any factor pair of 4096 corresponding to the timesteps and data dimension. The times of the data point in an EEG recording are crucial in identifying the type of patientbeing recorded. In the interest of exploring the time relation between data points, and improving our model's capacity of correctly classifying the two types of inter-ictal patients, different time steps and data dimensions were tested.

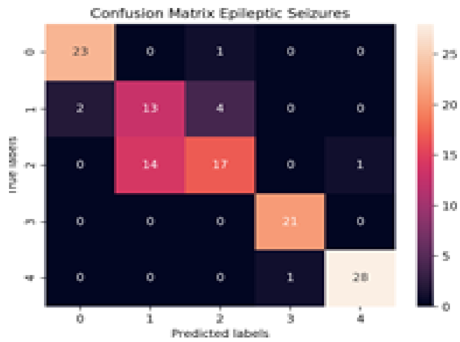

Figure 20 - confusion matrix 128 timesteps

Figure 20 shows the validation scores for different data dimensions (i.e number of recordings per timestep), averaged over 5 seeds; we can conclude that 256 timesteps (i.e 16 data dimensions per timestep) yield the best validation scores for our data (Figure 22, dim 16). Reducing the data dimension seems to have improved our model's generalization ability while overfitting less. This leads us to believe that shorter timesteps improves our model; this means our LSTM performs better on this dataset when less information is shared on between the neurons because shorter timesteps account for less variation in the recordings.

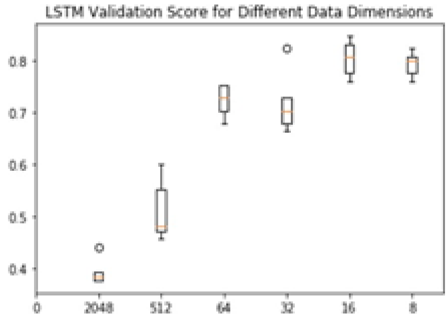

Figure 21 - results comparison timesteps

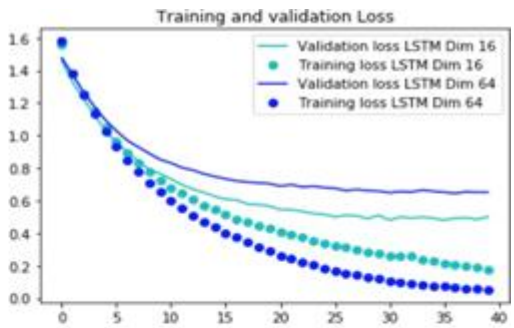

Figure 22 - training and validation loss

Figure 23 - Training and validation fitness

3.6 Neuron configurations

To change the configuration of the LSTM model in the dimension of the number of Neurons, two layers can be optimized. Firstly, the number of neurons in the LSTM layer, secondly the number of neurons in the hidden layer. Additionally, further layers can be added to the network, however in theory, this is not necessarilyrequired to solve this problem.

When adding neurons to the LSTM layer the accuracy increases. At 400 LSTM neurons the performance on validation accuracy is up around 5% compared to the model with only 100LSTM neurons. However, the runtime drastically increases morethan four-fold compared to 100 neurons in the first layer. Furthermore, an increase in overfitting is observable as the slope of the validation loss curve on the right in Figure 24 showsafter epoch 25.

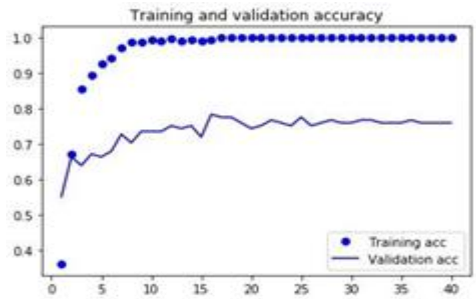

Figure 24 - training and validation fitness

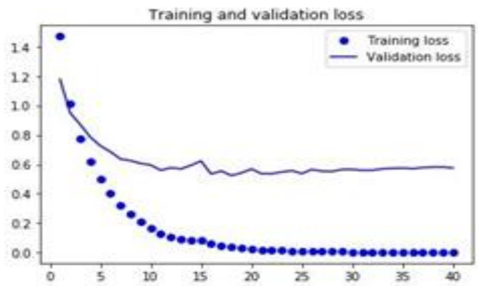

Figure 25 - training and validation loss

In other problems several concurrent LSTM layers, especially for text mining applications, proved useful. However, in our case this did not contribute to better results, probably because of thehigher complexity of text analysis.

When changing the number of neurons in the hidden layer, it seems the optimum number of neurons is at 50. Increasing or decreasing only weakens the performance of the classifier. An increase to over 100 neurons seems to have the same overfitting behavior as described above on the instance of the LSTM layer.

Adding additional hidden layers with different number of neurons did not contribute to better results, but either the results were worse or rather seed dependent, which we interpreted as another sign of overfitting.

3.7 Dropout & Regularization

Introducing Dropout and Regularization to the model are two measures to counter overfitting of the model. The gap between the training and validation data can possibly be reduced by applying these techniques more successfully.

Adding dropout after a layer of neurons of any kind will eliminate the weights at random positions of the model differently at every iteration of the model. Therefore, it will introduce a degree of randomness to the training of the weights.

Without any Dropout in the LSTM model introduced (black in figure 26 on the right), the model performs worst in terms of validation accuracy and validation loss compared to the blue boxes, representing one dropout layer after the LSTM and Dense layer each (dropout probability of 0.1 at optimum configuration), and the orange boxes representing the scores with only dropoutafter the LSTM layer.

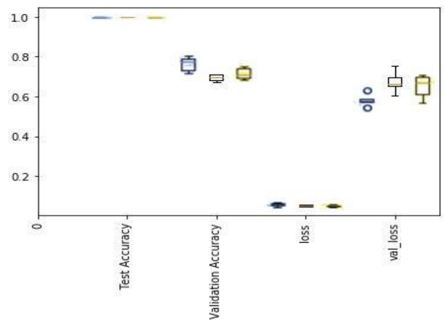

Figure 26 - results comparison dropout

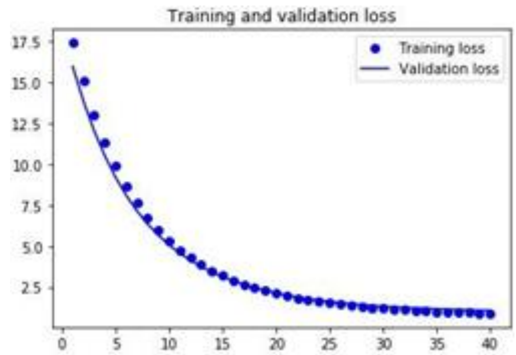

Figure 27 - regularized loss development

For the regularization, three regularizes are commonly used, the L1 (lasso) the L2 (ridge) and L1/L2 (elastic net) regularizes. All these regularizes can be appliedto the use of the activation inside the neuron, or to the output of the weights towards the loss function. However, in neither of the configurations, even with very low boundary values, did not improve the performance of the model or reduced overfitting. What can be observed though, is that the model learns a lot slower, and the initial error is much larger (10 fold) than withoutregularization as it is shown in figure 27 on the right.

4. CONCLUSION

For the five-class classification, the main objective of this paper, we noticed that our model had a tendency to over fit, Furthermore, while it correctly classified the three types of recordings (i.e seizure, inter-ictal, non-seizure), it had issues identifying the part of the brain being recorded for the inter-ictal patients. While this error is not dramatic in the sense that a wrong classification does not affect the health of a patient, a more accurate classifier would help researchers to better understand and treat patients. In order to tackle the issues of overfitting and wrong inter-ictal classification, different parameters and network architectures were explored. Reducing the dimension of the LSTM output through a Global Average Pooling did not improve our model. However, increasing the number of timesteps (i.e reducing the data dimensions) improved the model's generalization ability and slightly improved the model's classification ability; shorter timesteps lead to less recording variation being transferred between neurons in the LSTM layer, which leads to a better assessment of the recording at the previous time step. Finally, adding random neuron dropouts to the model, a known technique for combating overfitting, did not improve this particular model. To conclude,this LSTM model is capable of distinguishing seizure patients, inter-seizure patients, and healthy patients with high accuracy, making it a robust model with strong practical applications.

5. REFERENCES

[1] G. Rogers, “Epilepsy: the facts,” Primary Health Care Research & Development, vol. 11, no. 4, p. 413, 2010.

[2] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, “Automated EEG analysis of epilepsy: a review,” Knowledge-Based Systems, vol. 45, pp. 147–165, 2013.

[3] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, “Seizure prediction: the long and winding road,” Brain, vol. 130, no. 2, pp. 314– 333, 2006.

[4] R. Meier, H. Dittrich, A. Schulze-Bonhage, and A. Aertsen, “Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns,” Journal of Clinical Neurophysiology, vol. 25, no. 3, pp. 119– 131, 2008.

[5] G. R. Minasyan, J. B. Chatten, M. J. Chatten, and R. N. Harner, “Patientspecific early seizure detection from scalp EEG,” Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society, vol. 27, no. 3, p. 163, 2010.

[6] A. G. Correa, E. Laciar, H. Patino, and M. Valentinuzzi, “Artifact ˜ removal from EEG signals using adaptive filters in cascade,” in Journal of Physics: Conference Series, vol. 90, no. 1. IOP Publishing, 2007, p. 012081.

[7] A. M. Chan, F. T. Sun, E. H. Boto, and B. M. Wingeier, “Automated seizure onset detection for accurate onset time determination in intracranial EEG,” Clinical Neurophysiology, vol. 119, no. 12, pp. 2687–2696, 2008.

[8] A. Aarabi, R. Fazel-Rezai, and Y. Aghakhani, “A fuzzy rule-based system for epileptic seizure detection in intracranial EEG,” Clinical Neurophysiology, vol. 120, no. 9, pp. 1648–1657, 2009.

[9] J. J. Niederhauser, R. Esteller, J. Echauz, G. Vachtsevanos, and B. Litt, “Detection of seizure precursors from depth EEG using a sign periodogram transform,” IEEE Transactions on Biomedical Engineering, vol. 50, no. 4, pp. 449–458, 2003.

[10] I. Guler and E. D. ¨ Ubeyli, “Adaptive neuro-fuzzy inference system for ¨ classification of EEG signals using wavelet coefficients,” Journal of neuroscience methods, vol. 148, no. 2, pp. 113–121, 2005.

[11] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, “Automatic seizure detection based on time-frequency analysis and artificial neural networks,” Computational Intelligence and Neuroscience, vol. 2007, 2007.

[12] B. Abibullaev, H. D. Seo, and M. S. Kim, “Epileptic spike detection using continuous wavelet transforms and artificial neural networks,” International journal of wavelets, multiresolution and information processing, vol. 8, no. 01, pp. 33–48, 2010.

[13] J. Mitra, J. R. Glover, P. Y. Ktonas, A. T. Kumar, A. Mukherjee, N. B. Karayiannis, J. D. Frost Jr, R. A. Hrachovy, and E. M. Mizrahi, “A multi-stage system for the automated detection of epileptic seizures in neonatal EEG,” Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society, vol. 26, no. 4, p. 218, 2009.

[14] K. Abualsaud, M. Mahmuddin, M. Saleh, and A. Mohamed, “Ensemble classifier for epileptic seizure detection for imperfect EEG data,” The Scientific World Journal, vol. 2015, 2015.

[15] Zeng, Hong, Jiaming Zhang, Wael Zakaria, Fabio Babiloni, Borghini Gianluca, Xiufeng Li, and Wanzeng Kong. "InstanceEasyTL: an improved transfer-learning method for EEG-based cross-subject fatigue detection." *Sensors* 20, no. 24 (2020): 7251.

[16] Yıldırım, Özal, Ulas Baran Baloglu, and U. Rajendra Acharya. "A deep convolutional neural network model for automated identification of abnormal EEG signals." *Neural Computing and Applications* 32, no. 20 (2020): 15857-15868.

[17] RaviPrakash, Harish, Milena Korostenskaja, Eduardo M. Castillo, Ki H. Lee, Christine M. Salinas, James Baumgartner, Syed M. Anwar, Concetto Spampinato, and Ulas Bagci. "Deep Learning provides exceptional accuracy to ECoG-based Functional Language Mapping for epilepsy surgery." *Frontiers in neuroscience* (2020): 409.