II.PROPOSED WAVELET KERNEL INDUCED RESIDUAL NEURAL NETWORK (W-RESNET)

A Convolutional Neural Network (CNN/ConvNet) is an improvisation on the traditional ANN that replaces all but the last layer with specialised convolution layers. First introduced by Yann LeCun in 1997[1], CNNs have since seen widespread adoption in object detection, signal and image classification etc. But a commonly occurring drawback of deep CNNs was the problem of vanishing gradients. He et al. [2] in 2015 introduced the ResNet architecture, a ground-breaking work in the field of deep learning, which exhibited that extremely deep networks are trainable using the standard Stochastic Gradient Descent and a proper initialization function. ResNet is based on a micro-architecture called the Residual Module that relies on a series of operations used in usual CNNs including Batch Normalization, ReLU activation, Convolution and Max-pooling along with Identity mappings to combat the problem of vanishing gradients.

The ResNet model proposed here is a generalised, fully modular, customisable Residual neural network based on the ResNet V2 implementation with bottlenecking and pre activation. The architecture has been modified such that for the first convolutional layer, instead of the traditional Xavier normal or Xavier uniform initialisation function[3] a wavelet-based function has been used to initialise the kernel. The motivation behind the use of wavelets in this case was to discriminate between EEG signals obtained from healthy and epileptic individuals during both epileptic as well as seizure free interludes with eyes open and closed. Wavelet transforms [4] have the ability to describe a signal in both time and frequency domains as follows:

$$[W_\psi f](a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \overline{\psi\left(\frac{x-b}{a}\right)} f(x)dx \tag{1}$$

where,

$\overline{\psi\left(\frac{x-b}{a}\right)}$ is the selected mother wavelet, a and b being the scaling and shifting factors respectively and $f(x)$ is the input signal. Thus, wavelets have been used to extract the salient features of the different frequency bands of EEG signals.

Four mother wavelets were chosen to be used individually as well as together with varying degrees of performance. The number of wavelet convolution blocks used in parallel was increased from 1 to 4 and it was observed that combining better performing wavelet convolution blocks in parallel led to an increase in performance, whereas adding poorly performing wavelet convolution blocks decreased the performance of the neural network. Classical *Back propagation* algorithm has been used to update the kernel weights after each *epoch.* The schematic of the architecture of the proposed W-ResNet is depicted in fig. 1. The working principles of the proposed network is explained in the following subsections to provide a clear idea.

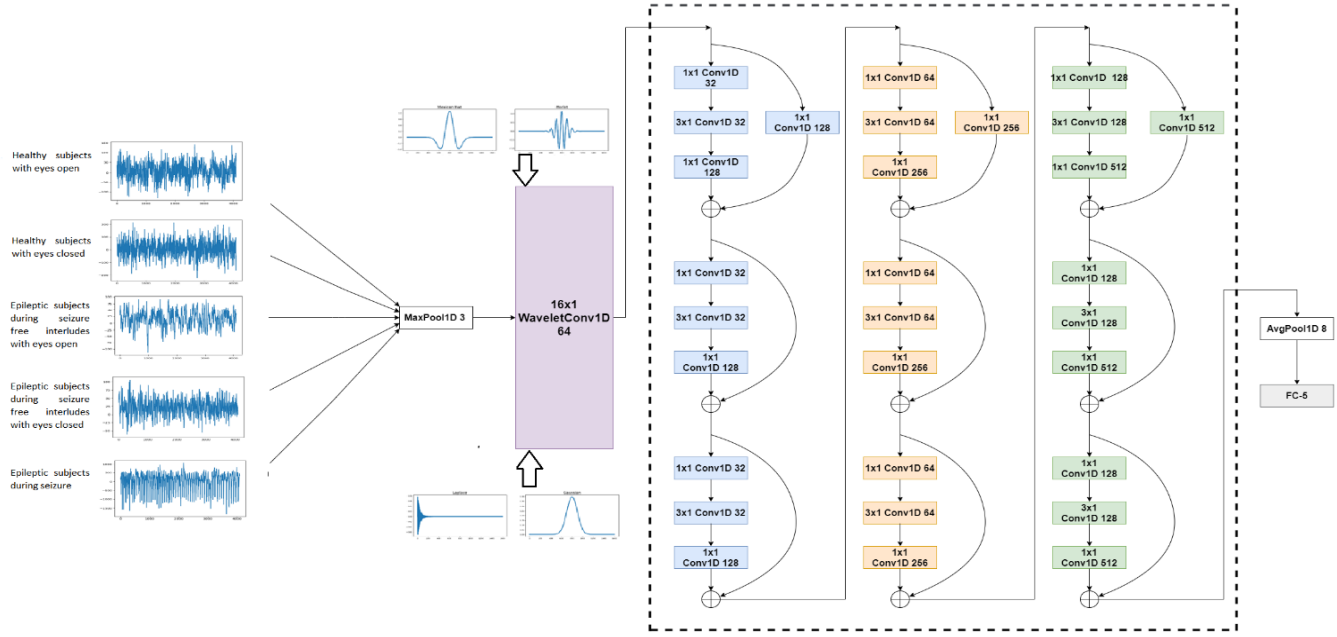

Fig. 1. W-ResNet Architecture

A. Input Layer (I)

This layer comprises of 5 classes of raw EEG data with a set of 100 samples available per class. Thus, a total of 500 samples are available. The raw EEG data have been fed to the W-ResNet. Each EEG data sample has a dimension of 1x4096.

B. Wavelet Kernel Based Convolution Layer (C)

This is an important layer of the proposed neural network architecture that primarily operates to extract the prominent and richer features. Four mother wavelets (Mexican Hat, Morlet, Laplace and Gaussian) have been chosen as the convolution kernels to extract salient time-frequency features of the input signals. The wavelet kernels used in the present proposition shown in fig. 2.

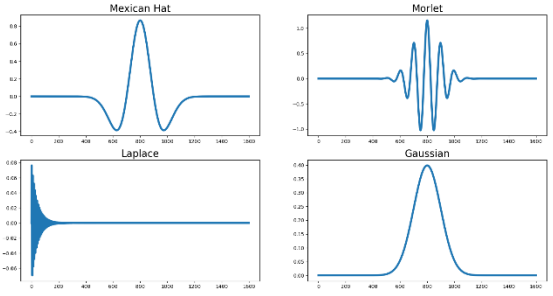

Fig. 2. Wavelet Kernels

are as follows:

- Mexican hat wavelet

$$\psi(t) = \frac{2}{\sqrt{3\sigma}\pi^{1/4}}\left(1-\left(\frac{t}{\sigma}\right)^2\right)e^{-\frac{t^2}{2\sigma^2}} \tag{2}$$

- Morlet wavelet

$$\Psi_\sigma(t) = c_\sigma \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2}\left(e^{i\sigma t} - \kappa_\sigma\right) \tag{3}$$

where,

$$\kappa_\sigma = e^{-\frac{1}{2}\sigma^2}$$

and

$$c_\sigma = \left(1 + e^{-\sigma^2} - 2e^{-\frac{3}{4}\sigma^2}\right)^{-\frac{1}{2}} \tag{4}$$

- Laplace wavelet

$$\psi_\zeta^{a,\tau} = \psi_\zeta\left(\frac{t-\tau}{a}\right) = \left(\frac{a^{-p}}{1-\zeta^2}\right)e^{\frac{-\zeta}{\sqrt{1-\zeta^2}}\left(\frac{t-\tau}{a}\right)} \times \sin\left(\frac{t-\tau}{a}\right)u(t-\tau) \tag{5}$$

where,

$$0 < \zeta < 1$$

specifies the wavelet family, τ is the time-shifting variable, a > 0 is the scaling variable, u(t) is the Heaviside step function, and p is a property preserving parameter, which takes the values of 1 or ½

- Gaussian wavelet

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (6)$$

The wavelet convolution blocks are implemented individually as well as together in parallel with varying performance metrics as follows:

$$C_i = \sum_{[j=1:4]} (S_i * W_j) \quad (7)$$

where, $S_{i\,(1 \le i \le 500)}$ is the raw EEG signal data, $W_{j\,(1 \le j \le 4)}$ is the raw EEG signal data, C_i is the convoluted signal and * is the convolution operator. The number of wavelet convolution blocks in parallel is increased from 1 up to 4. The wavelets with a size of (16×1) and a stride of 1, is chosen for the convolution operation.

Each convolution block has 64 output channels i.e., the wavelet filters are stacked to form a matrix of dimensions 64x16. The output of the convolution layer has a dimensionality of 4096x64. The convoluted signals are normalised using batch normalisation and passed through a ReLU (Rectified Linear Unit) activation function that suppresses all negative feature elements to 0 thereby improving the feature learning ability of the network.

$$C_i' = \mathrm{ReLU}\,(C_i) \quad (8)$$

Where ReLU activation function is defined as:

$$\mathrm{ReLU}(x) = x \text{ for } x \ge 0 \quad (9)$$

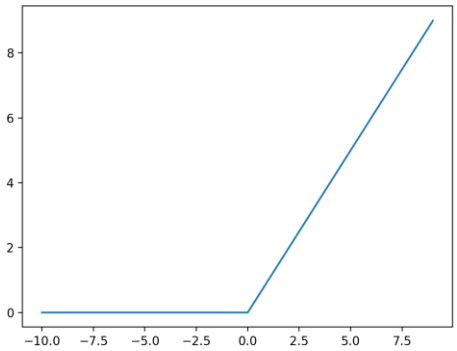

Fig. 3 ReLU function

C. Max Pooling Layer (M)

To reduce the dimension of the outputs of the activation function (Ci') are further processed through a zero-padding layer and a max-pooling layer. Max pooling performs down-sampling by dividing the signal into regions (pool size) and computes the maximum of each region.

$$M_i = max\,(C'_{i \times k:(i+2) \times k}) \quad (10)$$

In this layer, both the computational time and parameters are reduced. The pool size (k) for this layer is chosen as 3 and stride is chosen as 2. The dimension of the output signal (M_i) becomes 2048x64.

C. Residual Layer (R)

The specific implementation of ResNet that is used here consists of 29 layers comprising of 3 stages of each containing 3 residual blocks that learn 128, 256 and 512 filters in each stage, but the architecture is customisable and flexible to allow the user define the number of stages, the number of residual blocks per stage and the number of filters in each convolution and residual block. The model is a derivative of the ResNet v2 implementation by He et. al [2]. as well as the mxnet implementation by Wei Wu [5], but the parameters have been tuned for further efficiency.

The proposed ResNet implementation has 3 stages that stack 3, 3 and 3 residual modules each on convolution layers consisting of 128, 256 and 512 filters respectively. i.e.

In the first stage, 3 sets of residual modules are stacked such that the 3 convolution layers in each residual module learns 32, 32 and 128 convolution filters respectively.

The skip connections present in each Residual Module not only to transfer the information of the input signal, but also to produce an alternative path for gradient transformation and feature fusion. At the end of the stage spatial dimensions are reduced followed by batch normalisation to accelerate the learning process. The skip connection of the first residual module of each stage also has a convolution layer followed by a batch normalisation layer to reduce spatial dimensions

In the next stage, 3 more sets of residual modules are stacked such that each of the 3 convolution layers will learn 64, 64, and 256 filters. Afterwards, again a reduction is performed on the spatial dimensions followed by batch normalisation.

In the third stage, 3 more sets of residual modules are stacked, where each convolution layer of each residual module will learn 128, 128, and 512 filters respectively. a final reduction of spatial dimensions followed by batch normalisation is then carried out. The outputs of each Residual Module are expressed as:

$$R_i = (M_i * K_j) + M_i \quad (11)$$

where, M_i and R_i are the input and output of the Residual Module respectively.

The output is passed to a ReLU activation function

$$R_i' = \text{ReLU}(R_i) \quad (12)$$

D. Average Pooling Layer (A)

The dimensionality of the output of the activation function of the Residual layer is reduced by passing it to an Average pooling layer with a pool size of 8 and stride of 1

$$A_i = \text{avg}(R'_{i\times k:(i+1)\times k}) \quad (13)$$

E. Fully Connected Layer (FC)

In this layer each neuron is connected to the flattened outputs from the Average Pooling layer and a fully connected FC layer is formed with 5 neurons. Weights and biases from the Dense layer (FC) to output layer (O) are initialized using Xavier uniform initialisation function as detailed in [3], and the summed output is passed through SoftMax classifier to have the probability of the EEG signal class as a number in [0, 1] at the O. The value at the O is obtained as:

$$O = softmax(fc \circ w^T + b) \quad (14)$$

where, fc is the neuronal value at the FC layer, w is the weight matrix connected from FC to O, '$\circ$' is the element wise multiplication operator, b is the bias vector, and $softmax$ is the activation function, further defined as:

$$softmax = \frac{e^{x_p}}{\sum_{p=1}^{N} e^{x_p}} \quad (15)$$

where, x is the input to the $softmax$ function, p is the index and N is the number of neurons in the output layer.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Dataset Details

The Epilepsy dataset used for classification has been provided by University of Bonn [6]. It consists of 500 EEG signal data i.e., 5 cases with 100 samples per case. The 5 classes of data along with typical examples are as follows:

a. Healthy subjects with eyes open

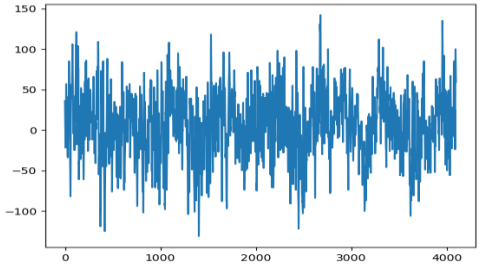

Fig. 4. Healthy subjects with eyes open

b. Healthy subjects with eyes closed

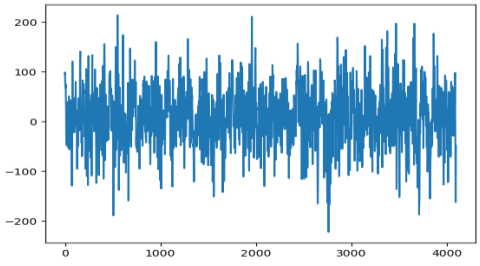

Fig. 5. Healthy subjects with eyes closed

c. Epileptic subjects during seizure free interludes with eyes open

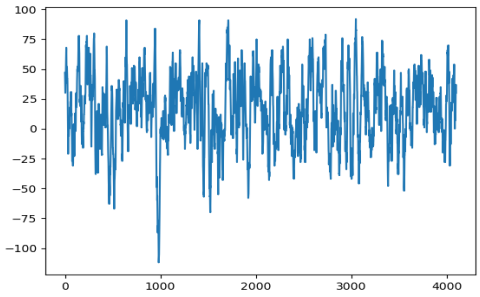

Fig. 6. Epileptic subjects during seizure free interludes with eyes open

d. Epileptic subjects during seizure free interludes with eyes closed

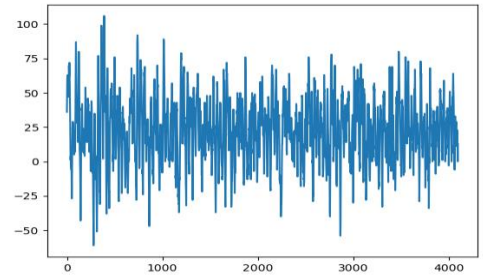

Fig. 7. Epileptic subjects during seizure free interludes with eyes closed

e. Epileptic subjects during seizure

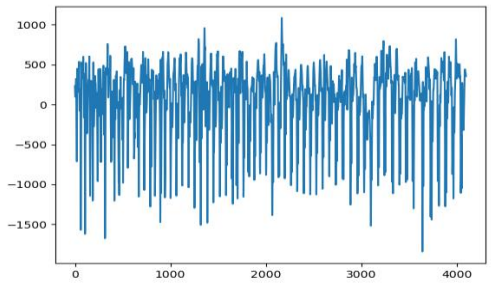

Fig. 8. Epileptic subjects during seizure

Of these, 300 (i.e., 60% of the total data) has been used for training ,100 (i.e., 20% of the total data) has been used for validation during training and 100 (i.e., 20% of the total data) has been used for testing the performance of the proposed neural network.

B. Training Details

The Deep Learning framework proposed here has been written completely in Anaconda distribution of the Python 3.7 programming language using the libraries TensorFlow and Keras and has a depth of 29 layers

This specific implementation of the W-ResNet architecture has been trained, validated and tested on the EEG dataset provided by University of Bonn [6] over an NVIDIA GTX 1050Ti GPU with 4GB of VRAM

Gradient descent learning has been used as the optimization function for the Back propagation algorithm to update the wavelet kernel weights with the minimization of the loss function. The loss function used is 'categorical cross-entropy'. The training epoch is set to 100 with an initial learning rate of 0.001 has been chosen with a momentum of 0.9. A polynomial function

$$\alpha = LR_{init} x (1-(e/maxEpochs))) \quad (16)$$

has been used as the learning rate scheduler

where, α is the learning rate for epoch e, LR_{init}

is the initial learning rate and maxEpochs is the total number of training epochs. A minibatch of size 20 has been chosen to train the proposed network. The residual modules of the proposed Resnet implementation takes a regularization strength of 0.0009, a batch normalisation epsilon of $23x10^{-6}$ (which is far lower than the Keras default of 0.001) but the batch normalisation momentum has been kept at 0.9 as recommended by He et. al [2]. and Wei Wu [5]. K-Fold Stratified cross validation has been applied for testing and training the model with the number of folds chosen as 4

i. Optimal Initial Learning Rate

The learning rate is a very important hyper-parameter for training any neural network. Until recently the optimal learning rate was chosen using a trial-and-error method. However, we have used the method proposed by Krishna Katyal on his GitHub project on Malaria Detection [7]. This method involves doing a trial run to train the neural network using a low learning rate, but then increasing it exponentially with each batch while recording the loss for every value of the learning rate. We then plot loss against learning rate. The optimum value of learning rate is the point where the learning rate is highest, and the slope of the loss vs learning rate curve is negative (i.e., loss is descending).

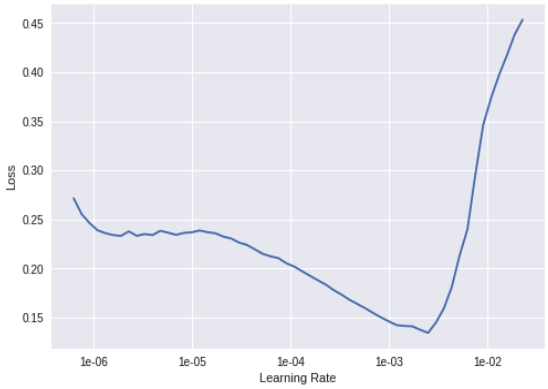

Fig.9.Loss vs Learning Rate curve

Based on the curve plotted above the optimised initial learning rate was chosen to be 0.001

ii. K-Fold Cross Validation

The K-fold cross validation is a statistical method used to estimate the skill of machine learning models. The method consists of a parameter k that represents the number of groups that a given data sample will be split into.

The general procedure is as follows:

The dataset is shuffled randomly and split into k groups. Each group is taken as the test set while the remaining k-1 groups are training set. The model is then fitted on training set and evaluated on test set. The evaluation score is retained while the model is discarded and the skill of the model is summarised using the retained evaluation scores. This procedure is repeated for all the groups.

Thus K-fold cross validation ensures that every observation from the original dataset has the chance of appearing in both training and test set and is a very useful technique for assessing model effectiveness, mitigating the overfitting problem and determining hyper parameters that will result in least test error. In this instance a 4-fold cross validation has been performed (i.e., k=4).

iii. Back Propagation in Wavelet Kernels

To calculate the error $\delta_{i,k}$ at the output layer the component wise difference of the target value and the output value is taken.

$$\delta_{i,k} = T_{i,k} - O_{i,k} \quad \forall k \tag{17}$$

where, $\delta_{i,k}$, $T_{i,k}$ and $O_{i,k}$ are the k-th component of i-th error vector, target vector and output vector respectively, and the cross-entropy loss function is defined as:

$$J = (-1/k) \sum_{[i=1:k]} \{T_i \,.\, \ln (O_i)\} \tag{18}$$

The error after each forward pass is back-propagated from the output layer to the previous layers by node updation of each layer followed by updation of the weights of the kernels in every layer. This procedure is known as the back-propagation (BP) algorithm and is based on the stochastic gradient descent technique. The BP process consists of the following two steps:

a) Node Updation

The errors calculated from each neuron of the Dense layer and propagated from the output layer to the input layer through the Dense layer

$$\delta_{fc} = \sum_{[i=1:N]} \delta_{i,k} \,.\, w \tag{19}$$

b) Weights and Biases Updation

Weights and biases are updated by the minimization of the cross-entropy loss function via gradient descent principle [6]

$$\left.\begin{aligned} w_{mn} &\leftarrow w_{mn} - \alpha \frac{\partial J}{\partial w_{mn}} \\ b_k &\leftarrow b_k - \alpha \frac{\partial J}{\partial b_k} \end{aligned}\right\} \tag{20}$$

where, w_{mn} is the weight connected between neuron m of (k-1) th layer to neuron n of k-th layer and α is the learning rate. The value of $\frac{\partial J}{\partial w_{mn}}$ is different for non-linear layers. From Output to the Dense layer, the non-linearity is SoftMax activation function. Hence, weights are updated as:

$$w_{mn} \leftarrow w_{mn} + \alpha(T_n - O_n).O_n\,(1 - O_n)\,O_m \tag{21}$$

where, O_n is the output obtained from n neuron at k-th layer and O_m is the output obtained from m neuron at (k – 1)-th layer. Since the max- pooling layers are linear, the weights of the wavelet kernels are updated as:

$$w_{mn} \leftarrow w_{mn} + \alpha\,(T_n - O_n).\,Om \tag{22}$$

Since, the weights are updated according to the BP algorithm and the updation rule is based on the gradient, there is a possibility of the network

being trapped in a *local minima*. In order to reduce this possibility, a momentum value of 0.9 is added and the weight adaptation rule becomes:

$$w_{mn} \leftarrow w_{mn} + \alpha(T_n - O_n).O_m + \gamma \Delta w_{mm} \quad (23)$$

where, γ is the momentum value and Δw_{mn} is the weight difference from node m to n. However, since there is no guarantee that the trapping to local minima can be completely avoided with this method, better initialization of the layer filters is a key factor in better learning or training of the network. By using wavelets that match the shape of the input signal in the first convolution layer, the kernels are better initialized. Therefore, they are more likely to reach the optimum value quicker (i.e., in lesser iterations) and achieve values for higher performance metrics like accuracy (ACC), specificity (SPE), sensitivity (SEN), positive predictive value (PPV) and negative predictive value (NPV). Similarly using wavelets that are not a good match for the input signals the kernels are poorly initialised resulting in higher loss and poor performance metrics as has been found in experimentation. Hence, proper selection of wavelet kernels and their combinations are necessary for better performance of the proposed neural network.

C. Results and Discussion

The proposed W-ResNet29 architecture has been used perform a multiclass classification analysis on the EEG dataset provided by University of Bonn comprising of five classes. The proposed W-ResNet architecture has been tested using all possible combinations of four different mother wavelets and the updation of the wavelet kernels throughout the epochs are displayed in fig.10.

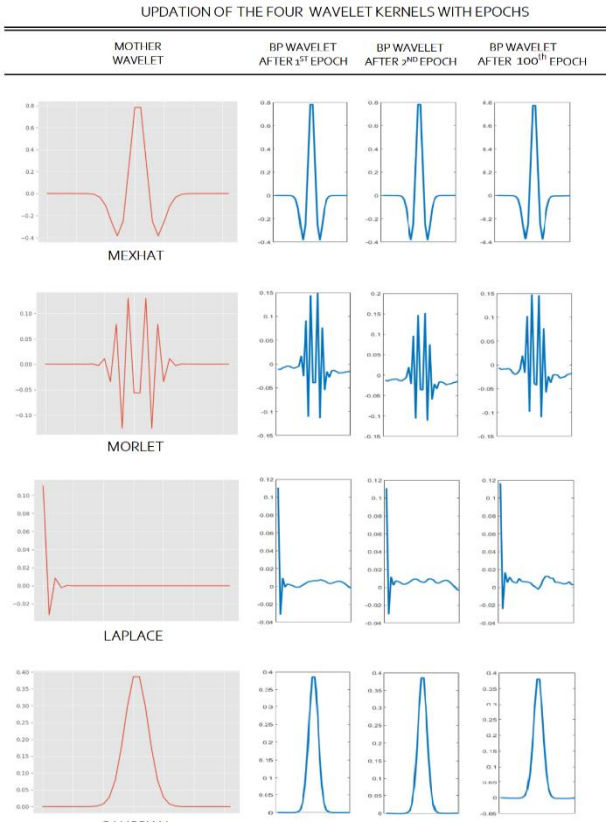

Fig.10. Updation of the 4 Wavelet Kernels with Epochs

Their performance metrics have also been compared with each other as well as with a classical ResNet29 architecture as presented in fig.11.

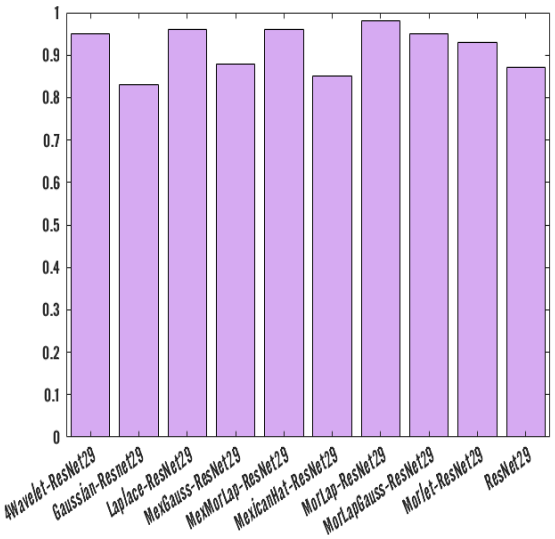

Fig. 11 Comparision of W-ResNet29 with different wavelets ,their combinations and standard ResNet29

The performances of all models have been evaluated using 4-fold cross-validation.

i. MexicanHat-ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass MexicanHat-ResNet29 has been shown in fig.10 and fig.11.

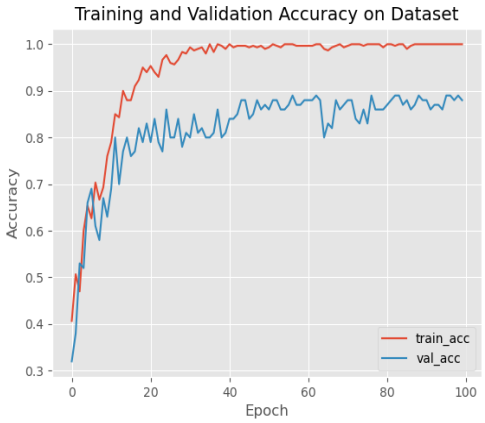

Fig. 10 Training and validation accuracy

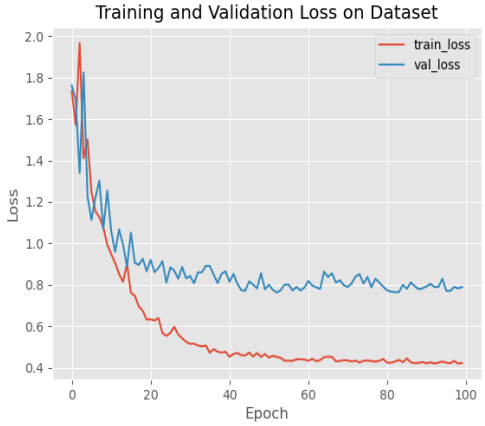

Fig. 11 Training and validation loss

The confusion matrix for this architecture is depicted in fig.12,

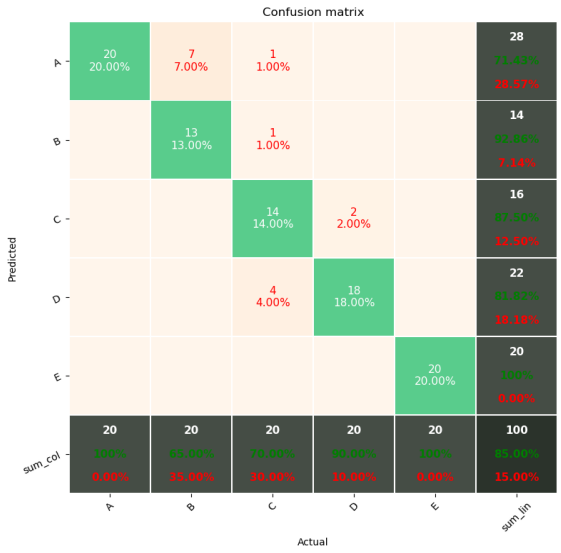

Fig. 12 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 85% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table I depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). We can see that by itself this kernel goes through negligible updations throughout epochs and as such performs poorly.

Table I Performance Metrics

	SPE	SEN	F1	SUPP
0	0.71	1.00	0.83	20
1	0.93	0.65	0.76	20
2	0.88	0.70	0.78	20
3	0.82	0.90	0.86	20
4	1.00	1.00	1.00	20
ACC			0.85	100
MACRO AVG	0.85	0.85	0.85	100
WEIGHTED AVG	0.85	0.85	0.85	100

ii. Morlet-ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass Morlet-ResNet29 has been shown in fig.13 and fig.14.

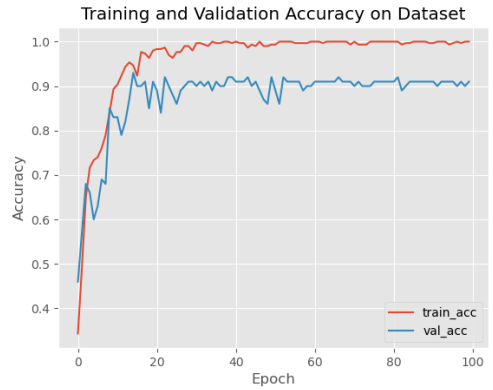

Fig. 13 Training and validation accuracy

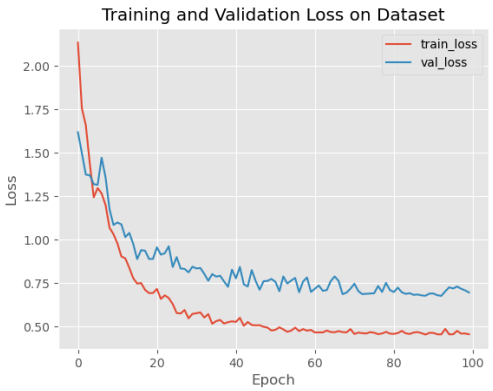

Fig. 14 Training and validation loss

The confusion matrix for this architecture is depicted in fig.15,

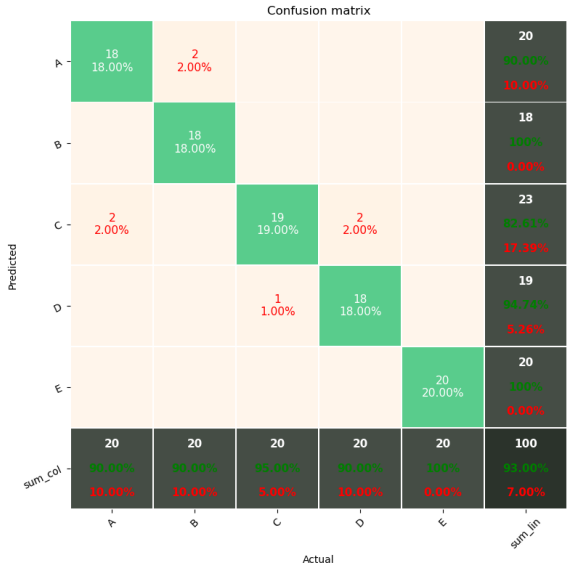

Fig. 15 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 93% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table II depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). This kernel shows a lot of adaptability and gives satisfactory performance.

Table II Performance Metrics

	SPE	SEN	F1	SUPP
0	0.90	0.90	0.90	20
1	1.00	0.90	0.95	20
2	0.83	0.95	0.88	20
3	0.95	0.90	0.92	20
4	1.00	1.00	1.00	20
ACC			0.93	100
MACRO AVG	0.93	0.93	0.93	100
WEIGHTED AVG	0.93	0.93	0.93	100

iii. Laplace-ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass Laplace-ResNet29 has been shown in fig.16 and fig.17.

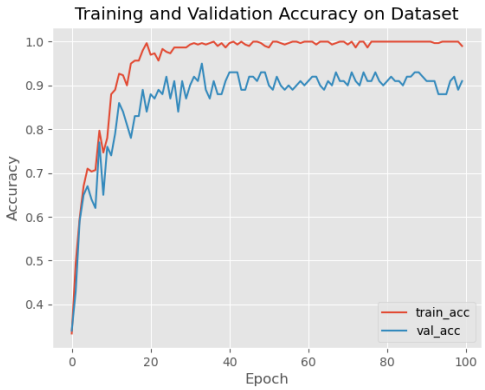

Fig. 16 Training and validation accuracy

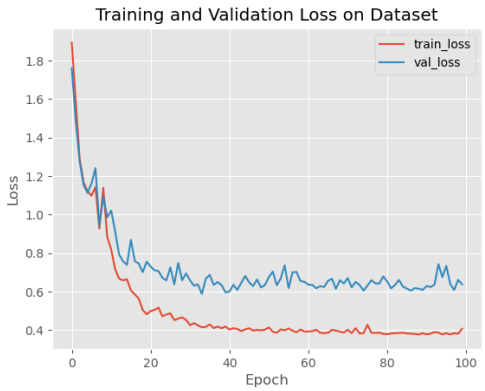

Fig. 17 Training and validation loss

The confusion matrix for this architecture is depicted in fig.18,

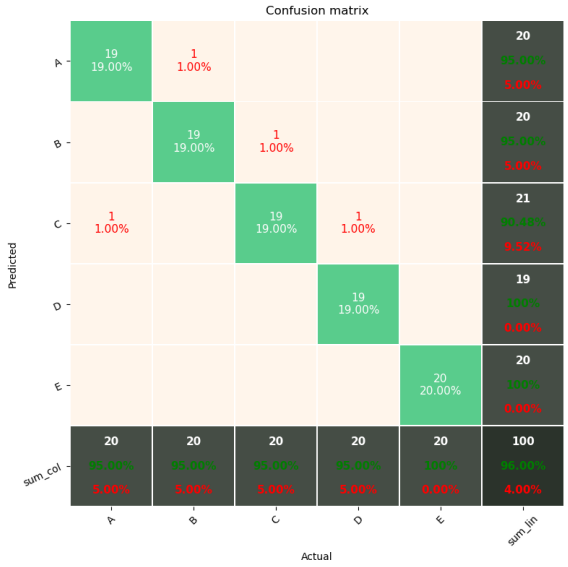

Fig. 18 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 96% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table III depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). This kernel has the highest adaptability and performs excellently by itself.

Table III Performance Metrics

	SPE	SEN	F1	SUPP
0	0.95	0.95	0.95	20
1	0.95	0.95	0.95	20
2	0.95	0.95	0.95	20
3	1.00	0.95	0.97	20
4	1.00	1.00	1.00	20
ACC			0.96	100
MACRO AVG	0.96	0.96	0.96	100
WEIGHTED AVG	0.96	0.96	0.96	100

iv. Gaussian-Resnet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass Gaussian-Resnet29 has been shown in fig.19 and fig.20.

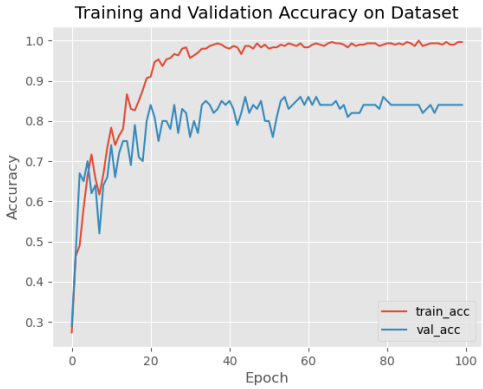

Fig. 19 Training and validation accuracy

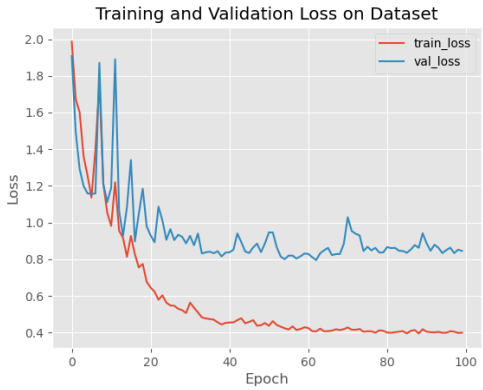

Fig. 20 Training and validation loss

The confusion matrix for this architecture is depicted in fig.21,

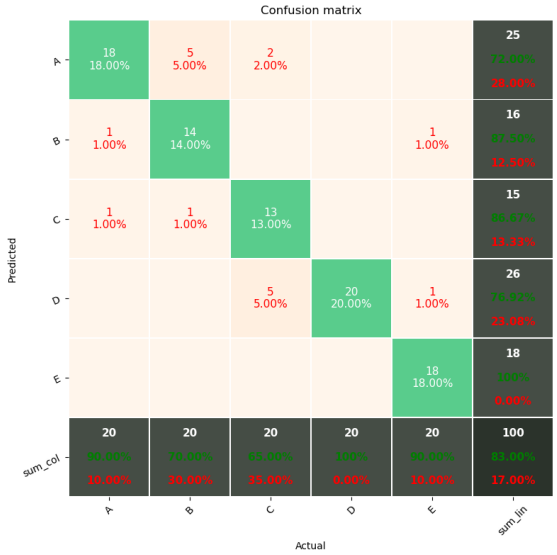

Fig. 21 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 83% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table IV depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). This kernel goes through the least updations and has the poorest performance

Table IV Performance Metrics

	SPE	SEN	F1	SUPP
0	0.72	0.90	0.80	20
1	0.88	0.70	0.78	20
2	0.87	0.65	0.74	20
3	0.77	1.00	0.87	20
4	1.00	0.90	0.95	20
ACC			0.83	100
MACRO AVG	0.83	0.83	0.83	100
WEIGHTED AVG	0.83	0.83	0.83	100

v. MexGauss-ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass MexGauss-ResNet29 has been shown in fig.22 and fig.23.

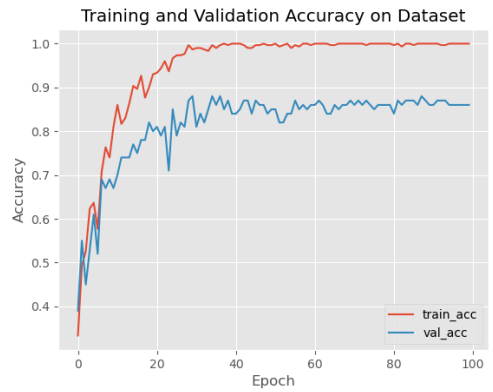

Fig. 22 Training and validation accuracy

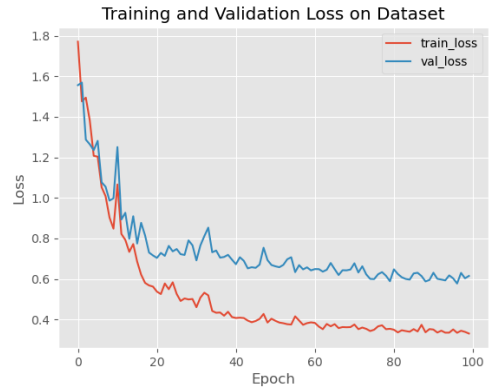

Fig. 23 Training and validation loss

The confusion matrix for this architecture is depicted in fig.24,

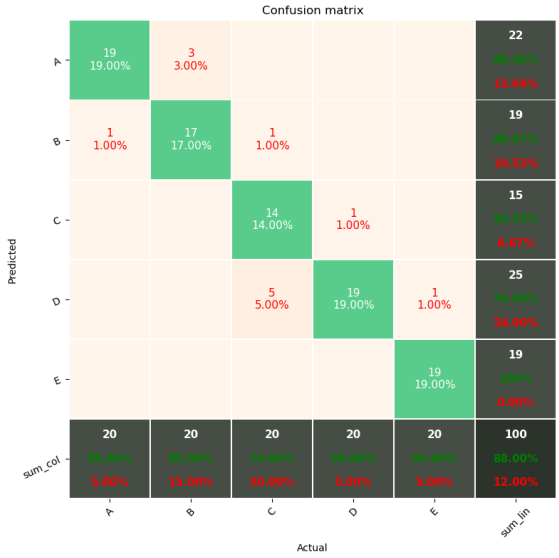

Fig. 24 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 88% is shown to have been obtained. The diagonal elements are an indication the correctly classified faults, whereas the off-diagonal elements show the misclassifications. Table V depicts the performance metrics of this EEG signals diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). Combining the 2 poorest performing kernels results in slightly better performance. However, the results are still not satisfactory.

Table V Performance Metrics

	SPE	SEN	F1	SUPP
0	0.86	0.95	0.90	20
1	0.89	0.85	0.87	20
2	0.93	0.70	0.80	20
3	0.76	0.95	0.84	20
4	1.00	0.95	0.97	20
ACC			0.88	100
MACRO AVG	0.88	0.88	0.88	100
WEIGHTED AVG	0.88	0.88	0.88	100

vi. MorLap-ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass MorLap-ResNet29 has been shown in fig.25 and fig.26. respectively.

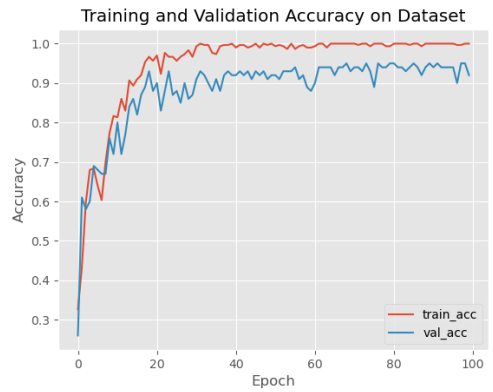

Fig. 25 Training and validation accuracy

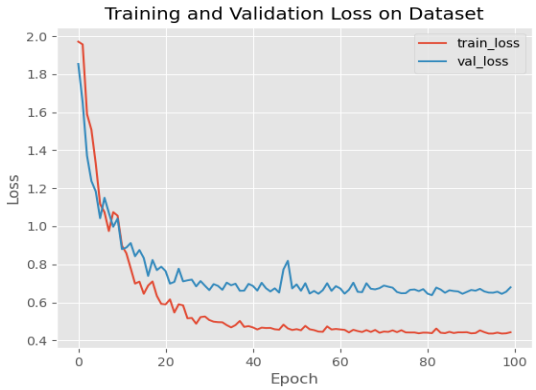

Fig. 26 Training and validation loss

The confusion matrix for this architecture is depicted in fig.27,

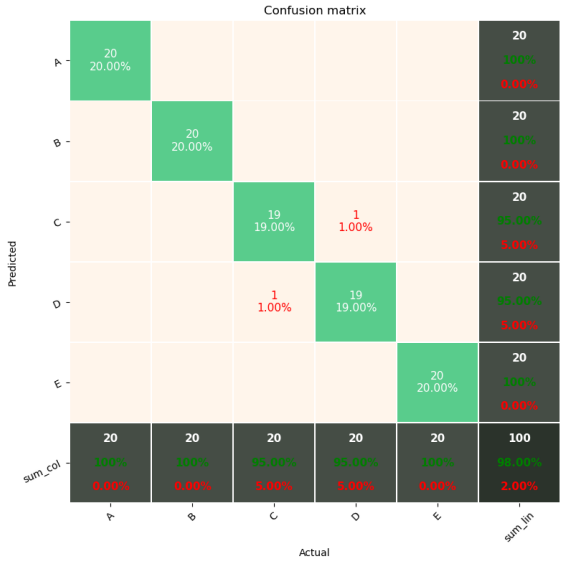

Fig. 27 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 98% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table VI depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). A combination of the best performing wavelets produces excellent results and is the best choice in this scenario.

Table VI Performance Metrics

	SPE	SEN	F1	SUPP
0	1.00	1.00	1.00	20
1	1.00	1.00	1.00	20
2	0.95	0.95	0.95	20
3	0.95	0.95	0.95	20
4	1.00	1.00	1.00	20
ACC			0.98	100
MACRO AVG	0.98	0.98	0.98	100
WEIGHTED AVG	0.98	0.98	0.98	100

vii. MexMorLap-ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass MexMorLap-ResNet29 has been shown in fig.28 and fig.29.

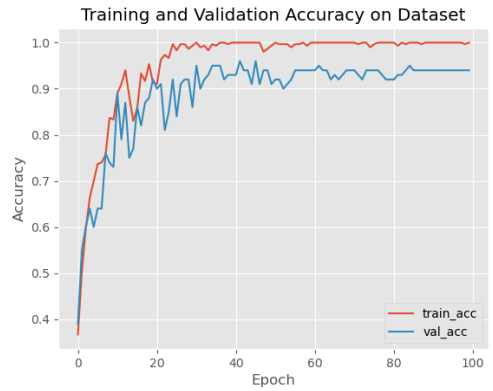

Fig. 28 Training and validation accuracy

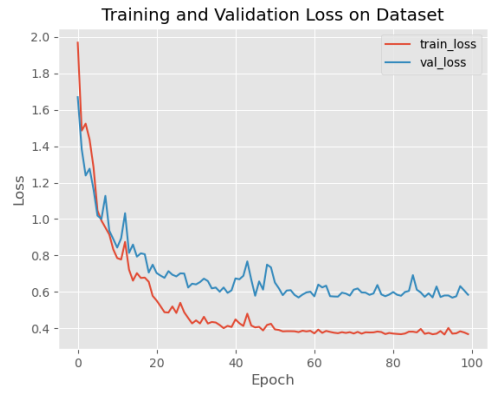

Fig. 29 Training and validation loss

The confusion matrix for this architecture is depicted in fig.30,

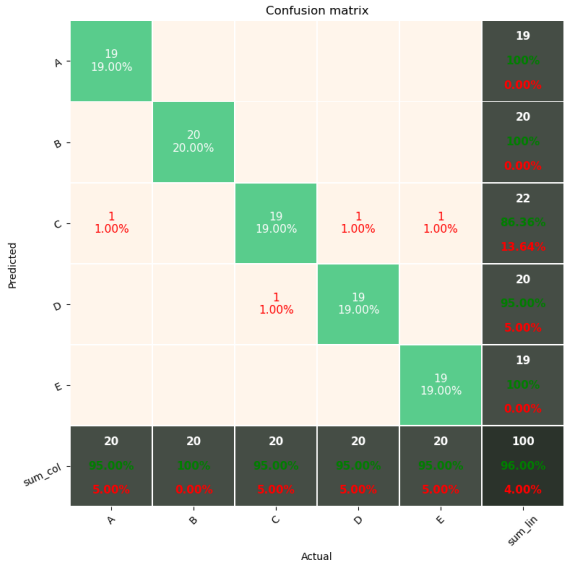

Fig. 30 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 96% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table VII depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). Addition of a poorly performing filter to the best performing filters slightly lowers their performance.

Table VII Performance Metrics

	SPE	SEN	F1	SUPP
0	1.00	0.95	0.97	20
1	1.00	1.00	1.00	20
2	0.86	0.95	0.90	20
3	0.95	0.95	0.95	20
4	1.00	0.95	0.97	20
ACC			0.96	100
MACRO AVG	0.96	0.96	0.96	100
WEIGHTED AVG	0.96	0.96	0.96	100

viii. MorLapGauss-ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass MorLapGauss-ResNet29 has been shown in fig.31 and fig.32.

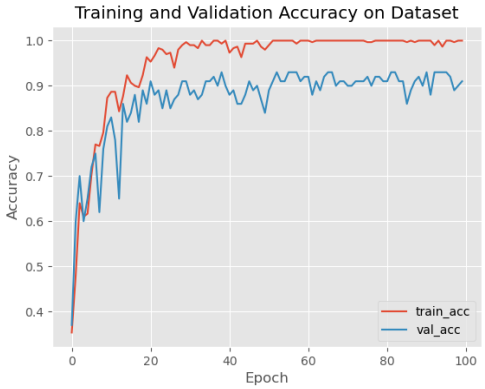

Fig. 31 Training and validation accuracy

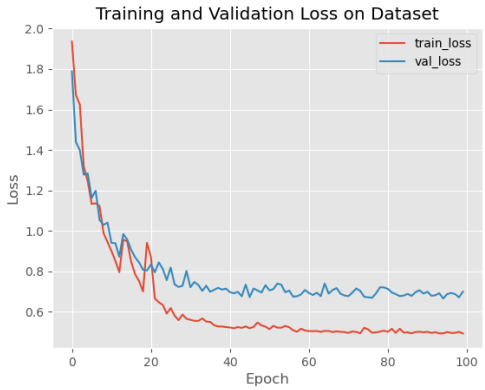

Fig. 32 Training and validation loss

The confusion matrix for this architecture is depicted in fig.33,

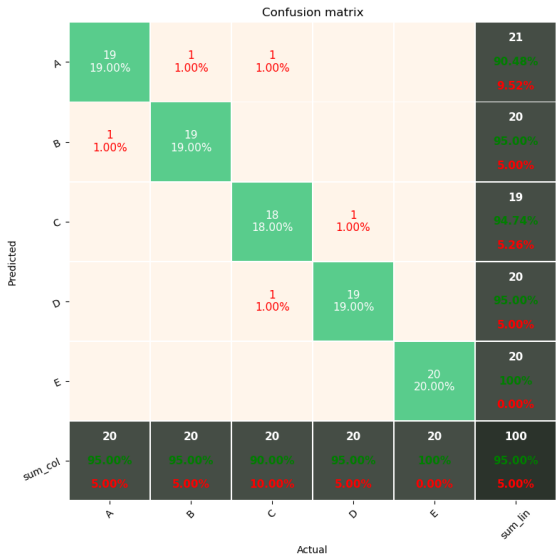

Fig. 33 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 95% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table VIII depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). Adding the poorest performing filter to the best performing wavelet filters further decreases their performance.

Table VIII Performance Metrics

	SPE	SEN	F1	SUPP
0	0.90	0.95	0.93	20
1	0.95	0.95	0.95	20
2	0.95	0.90	0.92	20
3	0.95	0.95	0.95	20
4	1.00	1.00	1.00	20
ACC			0.95	100
MACRO AVG	0.95	0.95	0.95	100
WEIGHTED AVG	0.95	0.95	0.95	100

ix. 4Wavelet-ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass 4Wavelet-ResNet29 has been shown in fig.34 and fig.35.

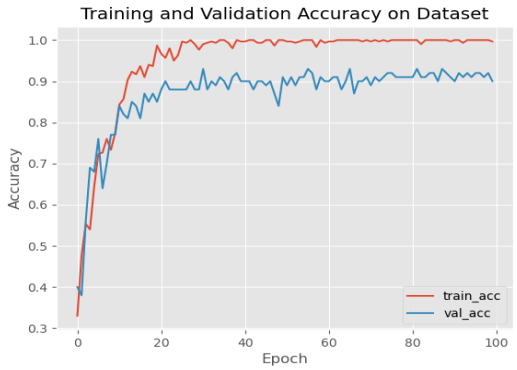

Fig. 34 Training and validation accuracy

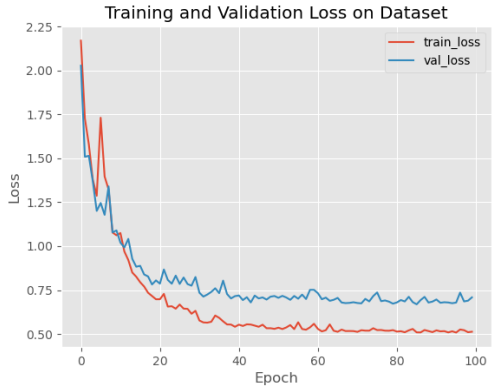

Fig. 35 Training and validation loss

The confusion matrix for this architecture is depicted in fig.36,

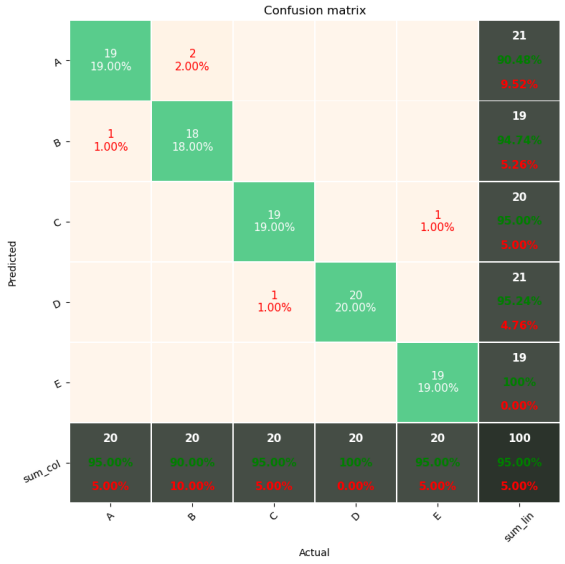

Fig. 36 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 95% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table IX depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). Combining all 4 mother wavelets has negligible effect on the performance .

Table IX Performance Metrics

	SPE	SEN	F1	SUPP
0	0.90	0.95	0.93	20
1	0.95	0.90	0.92	20
2	0.95	0.95	0.95	20
3	0.95	1.00	0.98	20
4	1.00	0.95	0.97	20
ACC			0.95	100
MACRO AVG	0.95	0.95	0.95	100
WEIGHTED AVG	0.95	0.95	0.95	100

x. Classical ResNet29

The accuracy and loss curves per epoch for training and validation for the proposed multiclass ResNet29 has been shown in fig.37 and fig.38.

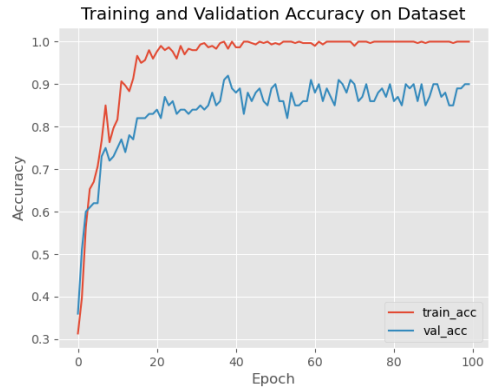

Fig. 37 Training and validation accuracy

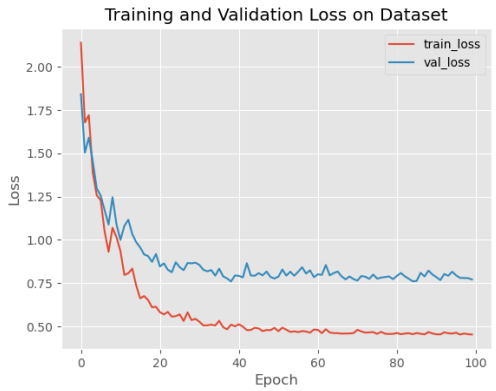

Fig. 38 Training and validation loss

The confusion matrix for this architecture is depicted in fig.39,

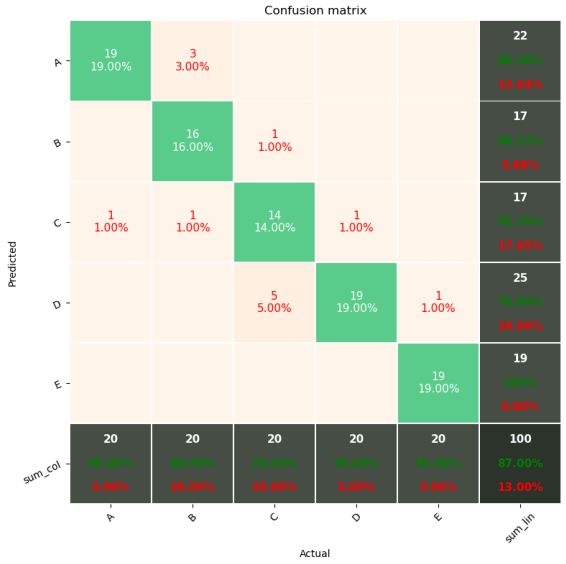

Fig. 39 Confusion Matrix

where, classes A, B, C, D and E each represent EEG signals obtained from Healthy subjects with eyes open, Healthy subjects with eyes closed, Epileptic subjects during seizure free interludes with eyes open, Epileptic subjects during seizure free interludes with eyes closed, Epileptic subjects during seizure respectively. An overall accuracy of 87% is shown to have been obtained. The diagonal elements are an indication the correctly classified signals, whereas the off-diagonal elements show the misclassifications. Table X depicts the performance metrics of this EEG signal diagnosis scheme by estimating four performance metrics, i.e., accuracy (ACC), specificity (SPE), sensitivity (SEN) and F1-score (F1). Standard ResNet with Xavier Normal [3] kernel gives a below satisfactory performance.

Table X Performance Metrics

	SPE	SEN	F1	SUPP
0	0.86	0.95	0.90	20
1	0.94	0.80	0.86	20
2	0.82	0.70	0.76	20
3	0.76	0.95	0.84	20
4	1.00	0.95	0.97	20
ACC			0.87	100
MACRO AVG	0.87	0.87	0.87	100
WEIGHTED AVG	0.87	0.87	0.87	100

IV. CONCLUSION

This paper presents a deep learning Neural Network architecture based on existing 1-D Residual Network using wavelet functions as kernel initialisers for automation of Epilepsy diagnosis from EEG signals. The superiority of the network lies in using multiple wavelet convolutions with skip connections present in the Residual modules for fusion of salient features, spatial reduction and activation of the last convolutional layer of the residual modules for accelerated learning. Smooth gradient transformation takes place in this architecture without suppression of any gradients resulting in lower computational time and an excellent accuracy of 98%, thereby outperforming state-of-the-art DL methods. In depth comparisons of the effects of different combinations of wavelet kernels have been provided and also compared with the performance of a standard Residual Network. It is seen that the best combination of mother wavelets outperforms the standard architecture by around 11%.

V. REFERENCES

[1] Y. Lecun and Y. Bengio, "Convolutional Networks for Images, Speech and Time Series",

[2] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition (2015)", ArXiv:1512.03385 [Cs], December, http://arxiv.org/abs/1512.03385

[3] X. Glorot, and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proc. 13th Int. Conf. Art. Intel. Stat. Sardinia, Italy, May 13-15, 2010, pp. 249-256.

[4] P.M. Bentley and J.T.E. McDonnell," Wavelet transforms: an introduction", Electronics & Communication Engineering Journal August 1994

[5] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T . Xiao, B. Xu, C .Zhang, Z. Zhang, W. Wu, "MX Net: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems", https://arxiv.org/abs/1512.01274v1

[6] Andrzejak RG, Lehnertz K, Rieke C, Mormann F, David P, Elger CE (2001) Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html

[7] K. Katyal, "Malaria Detection with Deep Learning", https://github.com/krishnakatyal/Malaria-Detection

[8] B. Ganguly, S. Chaudhuri, D. Dey, S. Munshi, B. Chatterjee, S. Dalai, S. Biswas, S. Chakravorti." Wavelet Kernel based Convolutional Neural Network for Localization of Partial Discharge Sources within a Power Apparatus"

[9] T. Abuhamdia, S. Taheri and J. Burns, "Laplace wavelet transform theory and applications", Journal of Vibration and Control 1–21

[10] T. Li, Z. Zhao, C. Sun, L. Cheng, X. Chen, R. Yan, and R. X. Gao," WaveletKernelNet: An Interpretable Deep Neural Network for Industrial Intelligent Diagnosis", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS