TemporalAugmenter: An Ensemble Recurrent Based Deep Learning Approach for Signal Classification

Nelly Elsayed
School of IT
University of Cincinnati
elsayeny@ucmail.uc.edu

Constantinos L. Zekios
Dep. of Elect. & Comp. Eng.
Florida International University
kzekios@fiu.edu

Navid Asadizanjani
Dep. of Elect. & Comp. Eng.
University of Florida
nasadi@ece.ufl.edu

Zag ElSayed
School of IT
University of Cincinnati
elsayezs@ucmail.uc.edu

Abstract

Ensemble modeling has been widely used to solve complex problems as it helps to improve overall performance and generalization. In this paper, we propose a novel TemporalAugmenter approach based on ensemble modeling for augmenting the temporal information capturing for long-term and short-term dependencies in data integration of two variations of recurrent neural networks in two learning streams to obtain the maximum possible temporal extraction. Thus, the proposed model augments the extraction of temporal dependencies. In addition, the proposed approach reduces the preprocessing and prior stages of feature extraction, which reduces the required energy to process the models built upon the proposed TemporalAugmenter approach, contributing towards green AI. Moreover, the proposed model can be simply integrated into various domains including industrial, medical, and human-computer interaction applications. Our proposed approach empirically evaluated the speech emotion recognition, electrocardiogram signal, and signal quality examination tasks as three different signals with varying complexity and different temporal dependency features.

Ensemble modeling is one of the solutions to overcome the model overfitting and improve the model performance by integrating multiple individual learning steams to create a robust and accurate predictive model, especially for complex tasks (Ganaie et al. 2022; Sagi and Rokach 2018). The ensemble modeling concept improves the overall model generalization, robustness, and stability and improves the overall model accuracy (Arpit et al. 2022; Ortega, Cabañas, and Masegosa 2022; Zhang, Cheng, and Hsieh 2019). Ensemble modeling has been applied in various applications including time series classification, speech recognition, image classification (Karim, Majumdar, and Darabi 2019; Kourentzes, Barrow, and Crone 2014; Elsayed, Maida, and Bayoumi 2018a), natural language processing (Sangamnerkar et al. 2020; Liu et al. 2019; Jia, Liang, and Liang 2023), events detection and recognition in videos (Adewopo and Elsayed 2023; Yu et al. 2020; Xu et al. 2018; Nanni et al. 2014), anomaly detection (Han, Chen, and Liu 2021; Zhao, Mehrotra, and Mohan 2015; Vanerio and Casas 2017), security of IoT devices (Tsogbaatar et al. 2021; Elsayed, ElSayed, and Bayoumi 2023; Alotaibi and Ilyas 2023), and medical applications (West et al. 2005; Priyadharshini et al. 2023; Liu et al. 2020).

Several ensemble modeling techniques in deep learning include boosting, bagging, stacking, negative correlation-based deep ensemble method, explicit and implicit ensembles, homogeneous and heterogeneous ensembles, and decision fusion strategies (Ganaie et al. 2022). The boosting technique is based on training the models sequentially, where each subsequent model focuses on solving the weaknesses of the previous model. Then, the highest weights are assigned to instances that were misclassified in the previous stages. That enables the model to learn from its mistakes, which helps to gradually improve the overall accuracy (Drucker et al. 1994; Ferreira and Figueiredo 2012; Mosavi et al. 2021). The bagging technique is based on training multiple models on subsets of the training data in parallel. Then, the final prediction is performed via aggregating all the models' predictions by taking a vote or the average (Altman and Krzywinski 2017; Galar et al. 2011). The stacking technique is based on combining multiple base model predictions by using an additional model called a meta-learner that is responsible for learning how to perform the best prediction based on the predictions of the multiple base models (Brownlee 2021). The negative correlation-based deep ensemble technique is based on training models that are negatively correlated by training the models in a way that aims to make different predictions, leading to promoting diverse predictions and reducing redundancy (Ganaie et al. 2022). The explicit technique combines multiple distinct models and performs the training explicitly (Ganaie et al. 2022). The implicit technique uses model uncertainty estimation within a single model or a dropout to create an implicit ensemble effect on the entire model (Seijo-Pardo et al. 2017). The homogeneous technique combines multiple models of the same type to enhance a prediction concept. The heterogeneous technique consists of different models in the concept to enhance the diverse learning strategy (Seijo-Pardo et al. 2017). The decision fusion technique is based on using multiple models to combine their final prediction decisions based on simple or complex methodology such as averaging, voting, or weights assigned to individual models's predictions (Ponti Jr 2011; Ganaie et al. 2022;

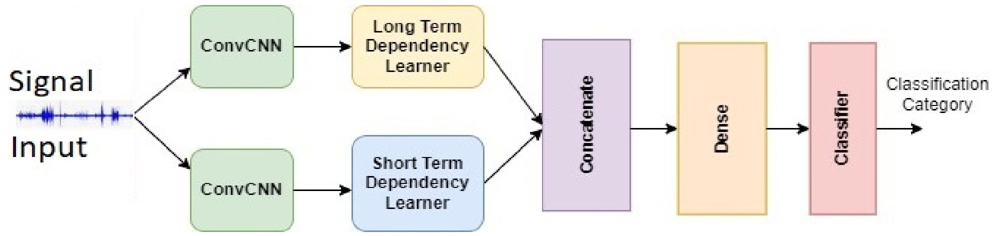

Figure 1: The proposed ensemble TemporalAugmenter approach for both long-term and short-term dependencies capturing in temporal data.

Hassan and Verma 2007).

There are different types of data depending on the primary source of data capturing, including discrete and sequential datasets (Dietterich 2002; Chmielewski and Grzymala-Busse 1996). The sequential data can be categorized into temporal data, where data points are collected, observed, and recorded at the same specific time intervals (e.g., videos, voice recording, time series, biological signal), and sequential (non-temporal) data that involves sequences where the order is significant. However, the observation time is not considered (e.g., text data, ordered events, and DNA sequences). The temporal data is complex due to the temporal information and the point-in-time information that the learning model must capture to perform the required task on the data. Thus, not all traditional learning models can solve temporal data based problems and tasks. Recurrent neural based architectures are the most suitable for capturing the temporal dependency information carried in the temporal data. There are several recurrent neural network based architectures such as the recurrent neural network (RNN) long short-term memory (LSTM) and its different variants (Greff et al. 2017; Elsayed, ElSayed, and Maida 2022; Gers, Schraudolph, and Schmidhuber 2002), the gated recurrent unit (GRU) and different variants (Chung et al. 2014; Dey and Salem 2017), the LiteLSTM (Elsayed, ElSayed, and Maida 2023), and the minimal gated unit MGU (Zhou et al. 2016). Each recurrent based network has its strengthes in capturing the long-term or short-term temporal dependencies. However, with the complex data, they require additional preprocessing or support for feature extracting to enhance the overall performance of capturing the point-to-time information.

Thus, in this paper, we propose a novel ensemble approach for augmenting the temporal information in temporal data based long-term and short-term dependencies capturing architectures: long short-term memory (LSTM) and the gated recurrent unit (GRU) in two streams that are capable of improving the overall performance. Moreover, we performed an empirical investigation of the influence of the convolutional neural network as a feature extractor prior to short-term temporal dependencies capturing architecture on improving the model performance; in addition, it eliminates the requirement of the data preprocessing stage. Finally, we employed the proposed approach on three different temporal tasks from different data sources and varied complexity, including speech emotion recognition, electrocardiogram classification, and radar signal quality classification, to analyze and validate the proposed approach concepts.

Temporal Augmenter Ensemble Approach

The proposed temporal augmenter approach is shown in Fig. 1 (Elsayed, ElSayed, and Maida 2022). The proposed approach consists of two stacked streams. The first stream employs a recurrent neural network architecture that is capable of extracting and learning the long-term temporal dependencies in the data. The second stream employs the recurrent neural network architecture that is capable of extracting and learning the short-term dependencies in the data. Adding one convolutional layer before each recurrent stream would improve the spatial features from the data, contributing to improving the overall approach performance at both temporal and spatial dependencies extraction and learning. Integrating the convolutional layer recurrent architectures has shown empirically and overall higher performance in several 1D applications including (Elsayed et al. 2022; Pan et al. 2020; Sajjad et al. 2020; Elsayed, ElSayed, and Bayoumi 2023). In this proposed approach, we empirically found that the optimal integration between the temporal dependencies extraction in the recurrent model and the convolution neural network (CNN) can be found while using the CNN as only one layer for extracting features prior to the recurrent network. Thus, the proposed approach reduces the required computations for the preprocessing of the signal due to the capability of the one-layer CNN to extract sufficient features, eliminating the requirements of signal preprocessing.

For the long-term dependency learning stream, in this approach, we employ the long short-term memory (LSTM) architecture as the main component for long-term dependency learning (Greff et al. 2016). The memory cell of the LSTM provided the capability of memorizing the long-term dependencies due to maintaining the long-term dependencies information in the learning stream. By reflecting the forget gate in the memory cell, the memory maintains the time dependencies that have long-term effects through the time in the memory, leading the LSTM to maintain a robust memorization of long-term dependencies from the data.

For the short-term dependencies learning stream, we selected the gated recurrent unit (GRU) as the primary component to learn the short-term dependencies in the data. The GRU is a smaller recurrent neural network architecture that consists of two gates: update z and reset gates r (Greff et al. 2017). The main concept was to share the weights between

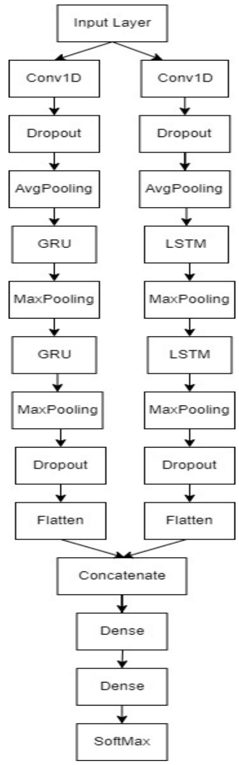

Figure 2: The proposed model based on the TemporalAugmenter appoach implementation layers and their orders.

two gates at the LSTM (input and forget) gates into one update gate and remove the memory cell to produce a smaller budget recurrent architecture that can be employed in applications where the sequences are small (e.g., short-term dependencies) (Chung et al. 2014). In addition, the GRU eliminates the output squashing function and the constant error carrocel (CEC) compared to the LSTM (Chung et al. 2014; Elsayed, ElSayed, and Maida 2023). The reset gate in the GRU maps the output gate of the LSTM. Thus, the GRU requires less budget to implement compared to the LSTM, and it is capable of learning short-term dependencies efficiently. Thus, the GRU has shown significant results and outperformed the LSTM in several applications where the time dependency in the data is short-term, such as (Shen et al. 2018; Elsayed, Maida, and Bayoumi 2019; Gao, Zheng, and Guo 2020; Yiğit et al. 2021; Elsayed, Zaghloul, and Li 2021; Golmohammadi et al. 2017; Wang et al. 2021; Elsayed, El-Sayed, and Bayoumi 2023; Jakubik 2018; Al-Shabandar et al. 2021).

Models that are based on our proposed TemporalAugmenter approach aim to employ both the LSTM as the long-term dependencies learning architecture with the GRU as a short-term dependencies learner into a model that is capable of capturing the long-term dependencies in temporal data as well as the short-term dependencies. Thus, the proposed TemporalAugmenter approach-based model can exceed the state-of-the-art models in multiple applications with different temporal data sources.

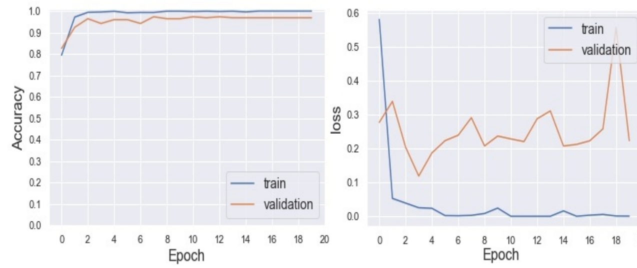

Figure 3: The proposed TemporalAugmenter approach training versus validation accuracy (left) and loss (right) diagrams over the TESS dataset.

Table 1: The proposed model overall statistics for speech emotion recognition from the TESS dataset.

Merits	Value
95% CI	(0.9783,0.9967)
Accuracy	99.64%
F1 Score	0.9875
False Negative Rate	0.0125
False Positive Rate	0.00208
True Negative Rate	0.99792
True Positve Rate	0.9875
Kappa	0.9854
Kappa 95% CI	(0.97465,0.99615)
Kappa Standard Error	0.00548
Total params	8,027,819
Trainable params	8,027,819
Non-trainable params	0

Experiments Setup

In our experiment, we targeted three different tasks: speech emotion recognition, electrocardiogram (ECG or EKG) classification, and radar signal quality classification tasks. These tasks are based on three different source of temporal data that vary in complexity and features behavior through time.

In our experiments, we implemented the models based on the proposed TemporalAugmenter approach. Figure 2 shows the model layers and order, which has been implemented based on the proposed TemporalAugmenter ensemble approach. The major differences between the experiments are in the number of epochs, batch size, optimization function, input size, and the number of classification categories. For the implementation, we used a computer with an Intel(R) Core(TM) i-9 CPU @ 3.00 GHz processor with 32-GB memory and NVIDIA GeForce RTX 2080 Ti graphics card. For the implementation, we used Tensorflow 2.4.0, Numpy 1.19.5, Pandas 1.2.4, Librosa 0.9.1, and Python 3.3.8 on a Windows 10 OS computer.

Task I: Speech Emotion Recognition

Dataset Description

In this task, we used audio-based speech emotion datasets, and we implemented the models to use the audio data di-

Table 2: The proposed model statistics over the seven speech emotion categories of the TESS dataset.

Statistical	Emotion Category						
Analysis	***Angry***	***Disgust***	***Fear***	***Happiness***	***Surprise***	***Sadness***	***Neutral***
Accuracy	100%	99.643%	99.464%	99.821%	99.643%	99.286%	99.643%
F1 Score	1.0	0.9878	0.97902	0.99355	0.9875	0.97826	0.98571
AUC	1.0	0.99286	0.97945	0.99359	0.99792	0.99131	0.99184
Error rate	0.0	0.00357	0.00536	0.00179	0.00357	0.00714	0.00357
False Negative Rate	0.0	0.0122	0.0411	0.01282	0.0	0.01099	0.01429
False Positive Rate	0.0	0.00209	0.0	0.0	0.00416	0.0064	0.00204
Specificity	1.0	0.99791	1.0	1.0	0.99584	0.9936	0.99796
Sensitivity	1.0	0.9878	0.9589	0.98718	1.0	0.98901	0.98571

Table 3: Comparison between the proposed model and the state-of-the-art models for speech emotion recognition over the TESS dataset.

Model	Method	Acc.
(Venkataramanan and Rajamohan 2019)	Combining 2D CNN and Global Avg. Pooling	66.00%
(Sundarprasad 2018)	Combining PCA, SVM, Mel-Frequeny, and Cepstrum Features	90.00%
(Krishnan, Joseph Raj, and Rajangam 2021)	SoA Classsifier and Entropy Features from Principle IMF modes	93.30%
(Lotfidereshgi and Gournay 2017)	Liquid State Machine	82.35%
(Zhang, Zhao, and Lei 2013)	Kernel Isomap	80.85%
(Zhang, Zhao, and Lei 2013)	PCA	72.35%
(Bhargava and Polzehl 2013)	Artificial Neural Nets	80.600%
(Bhargava and Polzehl 2013)	SVM	80.270%
(Elsayed et al. 2022)	1DCNN and GRU	94.285%
(Parry et al. 2019)	CNN and LSTM	49.48%
(Zhao, Mao, and Chen 2019)	2D-CNN and LSTM	70.00%
Our	**TemporalAugmenter**	**98.75%**

rectly without any mapping to spectrograms or assigning images or videos with the data. Thus, the proposed TemporalAugmenter approach aims to capture the temporal and spatial features from the spoken speech to determine the individual's emotions during the speech.

We used the Toronto Emotional Speech Set (TESS) dataset benchmark (Dupuis and Pichora-Fuller 2010). This data was recorded in the Toronto area by two actresses who have English as their first spoken language. This dataset consists of 2800 stimuli that represent seven different emotion categories. These emotions are anger, disgust, fear, happiness, surprise/pleasant, sadness, and neutral. The major advantage of this dataset is that the dataset is balanced between the number of stimuli among each of the seven classes.

Results and Analysis

In this experiment, the data was split to train, validate, and test with a ratio of 70%:10%:20%, respectively. The data has been scaled using standardscaler (Nabi and Nabi 2016). The short-term dependency stream 1D CNN that has 128 kernels of size one and the he_uniform function function as the kernel initializer. Then, the GRU number of units is set to 10, with glorot_uniform function as the kernel initializer. The long-term dependency stream started with a similar 1D CNN followed by the LSTM that has ten units, and the kernel initialization function is glorot_uniform and orthogonal function as the recurrent initializer. Then, the two streams concatenated, followed by two dense layers of 64 and 32, units with the rectified linear unit (ReLU) (Teh and Hinton 2000; Elsayed, Maida, and Bayoumi 2018b) as the activation functions. For training the model, the batch size has been set to 32 and the number of epochs to 20. RMSProp has been used as the optimization function with learning rate $lr = 0.001$, $momentum = 0.0$, and $\epsilon = 1e-07$. The categorical cross-

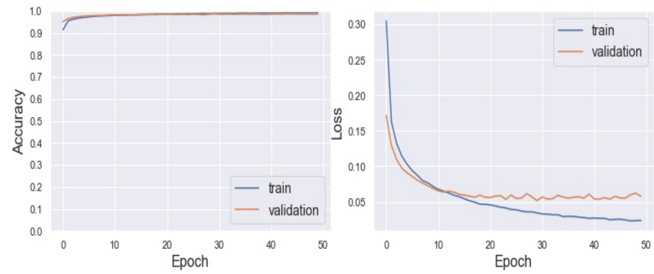

Figure 4: The proposed TemporalAugmenter approach training versus validation accuracy (left) and loss (right) diagrams over the MIT-BIH dataset.

Table 4: The proposed model overall statistics for MIT-BIH dataset.

Merits	Value
95% CI	(0.98293,0.98619)
Accuracy	98.456%
F1 Score	0.98456
False Negative Rate	0.01544
False Positive Rate	0.00208
True Negative Rate	0.99614
True Positve Rate	0.00386
Kappa	0.98456
Kappa 95% CI	(0.94317,0.95404)
Kappa Standard Error	0.00277
Total params	52,073
Trainable params	52,073
Non-trainable params	0

Table 5: The proposed model statistics over the five categories of the MIT-BIH dataset.

Statistical Analysis	ECG Category				
	N	*S*	*V*	*F*	*Q*
Accuracy	98.726%	99.287%	99.415%	99.694%	99.79%
F1 Score	0.99232	0.84942	0.95586	0.77888	0.9856
AUC	0.97164	0.89475	0.97698	0.86367	0.98913
Error rate	0.01274	0.00713	0.00585	0.00306	0.0021
False Negative Rate	0.00453	0.20863	0.04282	0.2716	0.02114
False Positive Rate	0.0522	0.00187	0.00323	0.00106	0.00059
Specificity	0.9478	0.99813	0.99677	0.99894	0.99941
Sensitivity	0.99547	0.79137	0.95718	0.7284	0.97886

Table 6: Comparison between the proposed model and the state-of-the-art models for MIT-BIH dataset.

Model	Method	Acc.
(Martis et al. 2013a)	DWT and SVM	93.8%
(Elsayed and Zaghloul 2020)	DWT and Random Forest	94.6%
(Asl, Setarehdan, and Mohebbi 2008)	DWT and LDA and RR	94.2%
(Osowski and Linh 2001)	Hybrid fuzzy NN	96.1%
(Acharya et al. 2017)	CNN and Augmentation	93.5%
(Kachuee, Fazeli, and Sarrafzadeh 2018)	Deep residual CNN	93.4%
(Elsayed and Zaghloul 2020)	ELM	96.4%
(Martis et al. 2013b)	SVM with RBF Kernel	93.5%
(Zhou, Jin, and Dong 2017)	CNN and LSTM	98.03%
(Acharya et al. 2017)	Daubechies Wavelet	94.3%
Our	**TemporalAugmenter**	**98.45%**

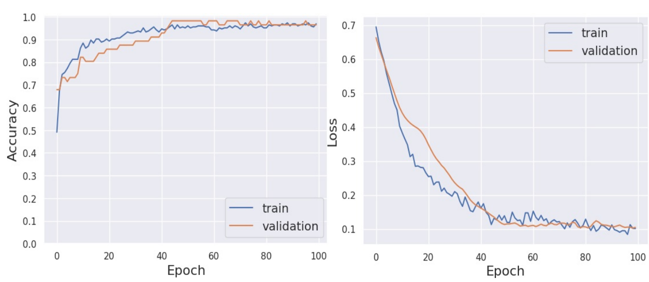

Figure 5: The proposed TemporalAugmenter approach training versus validation accuracy (left) and loss (right) diagrams over the Radar Ionosphere Depletion dataset.

entropy is set as the loss function. Max pooling and dropouts have been applied in both the long-term and short-term dependency streams. Fig. 3 shows the proposed model training versus validation accuracy and loss. Table 1 shows the proposed model's overall statistics over the TESS dataset. Table 2 shows the statistics of the model over each of the seven emotions, where (OP) and (AUC) are the areas under the Receiver Operating Characteristic (ROC) curve. Table 3 compares the proposed model and the state-of-the-art speech emotion recognition models over the TESS dataset.

Table 7: The proposed model overall statistics for the Radar Ionosphere Depletion dataset.

Merits	Value
95% CI	(0.91095, 1.0)
Accuracy	95.775%
F1 Score	0.95775
False Negative Rate	0.04225
False Positive Rate	0.04225
True Negative Rate	0.95775
True Positve Rate	0.95775
Kappa	0.90839
Kappa 95% CI	(0.80693,1.00)
Kappa Standard Error	0.05176
Total params	21,214
Trainable params	21,214
Non-trainable params	0

Table 8: The proposed model statistics over the two categories of the Radar Ionosphere Depletion dataset.

Statistical Analysis	Radar Signal Category	
	Class 0	*Class 1*
Accuracy	0.95775	0.95775
F1 Score	0.96703	0.94118
AUC	0.94444	0.94444
Error rate	0.04225	0.04225
False Negative Rate	0.0	0.11111
False Positive Rate	0.11111	0.0
Specificity	0.88889	1.0
Sensitivity	1.0	0.88889

Task II: Electrocardiogram Classification

Dataset Description

In this task, we aim to empirically evaluate the TemporalAugmenter concept of the electrocardiogram (ECG) classification task as an example of a biological signal that carries information about heart functionality. In this experiment, we used the MIT-BIH dataset benchmark collected by the BIH Laboratory (Moody and Mark 2001). The ECG signals were recorded from 25 women between 32 to 89 years old and 22 women aged between 23 to 89 years old. The dataset consists of 109,446 data samples. The dataset consists of five different categories of the ECG recorded signal,

Table 9: Comparison between the proposed model and the state-of-the-art models for Radar Ionosphere Depletion dataset.

Model	Method	Acc.
(Basheer et al. 2024)	AADS	76.13%
(Basheer et al. 2024)	Streaming TEDA	52.38%
(Basheer et al. 2024)	MAD	22.35%
(Basheer et al. 2024)	xStream	21.38%
(Basheer et al. 2024)	RRCF	12.65%
(Sigillito et al. 1989)	linear perceptron	90.67%
(Sigillito et al. 1989)	MLFN	83.8%
Our	**TemporalAugmenter**	**95.775%**

including normal heart beat (N), supraventricular premature beat (S), premature ventricular contraction (V), fusion of paced and normal beat (F), and unclassifiable beat (Q).

Results and Analysis

For the experiment, we divided into training, validation, and testing by the ratio 60%, 20%, and 20%, respectively. The batch size is set to 128, and the number of epochs to 50. The first ratio of dropout layers at the long-term and short-term steams was set to 50%, maintaining the rest at the 30% ratio. Adam optimizer has been used with learning rate $lr = 0.001$, and $\epsilon = 1e-07$. The proposed model training versus validation accuracy and loss diagrams are shown in Figure 4. Table 4 shows the proposed model's overall statistics over the MIT-BIH dataset. Table 5 shows the statistics of the model over each of the five ECG categories. Table 6 compares the proposed model and the state-of-the-art classification models over the MIT-BIH dataset.

Task III: Radar Signal Quality Classification

Dataset Description

In this experiment, we used the Ionosphere Depletion dataset benchmark collected by a Goose Bay system to evaluate the Ionosphere. The dataset consists of 351 data samples. The pulse numbers of the Good Bay system were 17. Each data sample in the dataset is described by two attributes per pulse number, corresponding to the complex values returned by the function resulting from the complex electromagnetic signal. Thus, the total number of attributes is 34. The data consists of two categories, bad signal and good signal, which are decoded to zero and one in the implementation. The dataset is unbalanced.

Results and Analysis

In this experiment, we split the data to train, validate, and test the dataset with ratios 60%, 20%, and 20%, respectively. The Adam optimizer has been used with a learning rate $lr = 0.001$, and $\epsilon = 1e-07$. The number of epochs is set to 100, and the batch size to 128. The proposed model overall statistics for the Radar Ionosphere Depletion dataset experiment are shown in Table 7. The proposed model statistics over the two categories of the Radar Ionosphere Depletion dataset are shown in Table 8. Table 9 compares the proposed model and the state-of-the-art Radar Ionosphere Depletion dataset classification models. Our proposed model exceeds the state-of-the-art models' accuracy performance.

Conclusion and Social Impact

Manipulating temporal data requires robust methodologies that can capture both temporal and point-to-time information. Several applications are based on classifying temporal data, such as biosignals classification for diagnostics, speech emotion recognition, stock market prediction, energy consumption, time series anomaly detection, and radar signal classification. This paper proposed a simple ensemble based approach, TemporalAugmenter, based on integrating long-term and short-term dependency learning streams to capture precise temporal dependencies in temporal data. In addition, we found empirically that one convolutional layer prior to a recurrent architecture can help enhance the model performance. Furthermore, the proposed approach can be used directly without extracting additional features, preprocessing, or converting the signal to spectrograms, which reduces the power and energy required for model implementation, contributing towards green AI and reducing CO_2 footprint. Thus, models that are based on the proposed TemporalAugmenter approach can be implemented within different temporal data based application domains.

References

Acharya, U. R.; Oh, S. L.; Hagiwara, Y.; Tan, J. H.; Adam, M.; Gertych, A.; and San Tan, R. 2017. A deep convolutional neural network model to classify heartbeats. *Computers in biology and medicine* 89:389–396.

Adewopo, V., and Elsayed, N. 2023. Smart city transportation: Deep learning ensemble approach for traffic accident detection. *arXiv preprint arXiv:2310.10038*.

Al-Shabandar, R.; Jaddoa, A.; Liatsis, P.; and Hussain, A. J. 2021. A deep gated recurrent neural network for petroleum production forecasting. *Machine Learning with Applications* 3:100013.

Alotaibi, Y., and Ilyas, M. 2023. Ensemble-learning framework for intrusion detection to enhance internet of things' devices security. *Sensors* 23(12):5568.

Altman, N., and Krzywinski, M. 2017. Ensemble methods: bagging and random forests. *Nature Methods* 14(10):933–935.

Arpit, D.; Wang, H.; Zhou, Y.; and Xiong, C. 2022. Ensemble of averages: Improving model selection and boosting performance in domain generalization. *Advances in Neural Information Processing Systems* 35:8265–8277.

Asl, B. M.; Setarehdan, S. K.; and Mohebbi, M. 2008. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial intelligence in medicine* 44(1):51–64.

Basheer, M. Y. I.; Ali, A. M.; Hamid, N. H. A.; Ariffin, M. A. M.; Osman, R.; Nordin, S.; and Gu, X. 2024. Autonomous anomaly detection for streaming data. *Knowledge-Based Systems* 284:111235.

Bhargava, M., and Polzehl, T. 2013. Improving automatic emotion recognition from speech using rhythm and temporal feature. *arXiv preprint arXiv:1303.1761*.

Brownlee, J. 2021. *Ensemble learning algorithms with Python: Make better predictions with bagging, boosting, and stacking*. Machine Learning Mastery.

Chmielewski, M. R., and Grzymala-Busse, J. W. 1996. Global discretization of continuous attributes as preprocessing for machine learning. *International journal of approximate reasoning* 15(4):319–331.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Dey, R., and Salem, F. M. 2017. Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, 1597–1600. IEEE.

Dietterich, T. G. 2002. Machine learning for sequential data: A review. In *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops SSPR 2002 and SPR 2002 Windsor, Ontario, Canada, August 6–9, 2002 Proceedings*, 15–30. Springer.

Drucker, H.; Cortes, C.; Jackel, L. D.; LeCun, Y.; and Vapnik, V. 1994. Boosting and other ensemble methods. *Neural computation* 6(6):1289–1301.

Dupuis, K., and Pichora-Fuller, M. K. 2010. Toronto emotional speech set (TESS)-younger talker_happy.

Elsayed, N., and Zaghloul, Z. S. 2020. A simple extreme learning machine model for detecting heart arrhythmia in the electrocardiography signal. In *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*, 513–516. IEEE.

Elsayed, N.; ElSayed, Z.; Asadizanjani, N.; Ozer, M.; Abdelgawad, A.; and Bayoumi, M. 2022. Speech emotion recognition using supervised deep recurrent system for mental health monitoring. In *2022 IEEE 8th World Forum on Internet of Things (WF-IoT)*, 1–6. IEEE.

Elsayed, N.; ElSayed, Z.; and Bayoumi, M. 2023. Iot botnet detection using an economic deep learning model. *arXiv preprint arXiv:2302.02013*.

Elsayed, N.; ElSayed, Z.; and Maida, A. S. 2022. Litelstm architecture for deep recurrent neural networks. In *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1304–1308. IEEE.

Elsayed, N.; ElSayed, Z.; and Maida, A. S. 2023. Litelstm architecture based on weights sharing for recurrent neural networks. *arXiv preprint arXiv:2301.04794*.

Elsayed, N.; Maida, A. S.; and Bayoumi, M. 2018a. Deep gated recurrent and convolutional network hybrid model for univariate time series classification. *arXiv preprint arXiv:1812.07683*.

Elsayed, N.; Maida, A. S.; and Bayoumi, M. 2018b. Empirical activation function effects on unsupervised convolutional lstm learning. In *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, 336–343. IEEE.

Elsayed, N.; Maida, A. S.; and Bayoumi, M. 2019. Gated recurrent neural networks empirical utilization for time series classification. In *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 1207–1210. IEEE.

Elsayed, N.; Zaghloul, Z. S.; and Li, C. 2021. Arrhythmia supraventricular premature beat detection in electrocardiography signal using deep gated recurrent model. In *SoutheastCon 2021*, 1–7. IEEE.

Ferreira, A. J., and Figueiredo, M. A. 2012. Boosting algorithms: A review of methods, theory, and applications. *Ensemble machine learning: Methods and applications* 35–85.

Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; and Herrera, F. 2011. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42(4):463–484.

Ganaie, M. A.; Hu, M.; Malik, A.; Tanveer, M.; and Suganthan, P. 2022. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence* 115:105151.

Gao, S.; Zheng, Y.; and Guo, X. 2020. Gated recurrent unit-based heart sound analysis for heart failure screening. *Biomedical engineering online* 19:1–17.

Gers, F. A.; Schraudolph, N. N.; and Schmidhuber, J. 2002. Learning precise timing with lstm recurrent networks. *Journal of machine learning research* 3(Aug):115–143.

Golmohammadi, M.; Ziyabari, S.; Shah, V.; Von Weltin, E.; Campbell, C.; Obeid, I.; and Picone, J. 2017. Gated recurrent networks for seizure detection. In *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 1–5. IEEE.

Greff, K.; Srivastava, R. K.; Koutník, J.; Steunebrink, B. R.; and Schmidhuber, J. 2016. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems* 28(10):2222–2232.

Greff, K.; Srivastava, R. K.; Koutník, J.; Steunebrink, B. R.; and Schmidhuber, J. 2017. LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems* 28(10):2222–2232.

Han, X.; Chen, X.; and Liu, L.-P. 2021. Gan ensemble for anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 4090–4097.

Hassan, S. Z., and Verma, B. 2007. Decisions fusion strategy: Towards hybrid cluster ensemble. In *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, 377–382. IEEE.

Jakubik, J. 2018. Evaluation of gated recurrent neural networks in music classification tasks. In *Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology–ISAT 2017: Part I*, 27–37. Springer.

Jia, J.; Liang, W.; and Liang, Y. 2023. A review of hybrid and ensemble in deep learning for natural language processing. *arXiv preprint arXiv:2312.05589*.

Kachuee, M.; Fazeli, S.; and Sarrafzadeh, M. 2018. Ecg heartbeat classification: A deep transferable representation. In *2018 IEEE international conference on healthcare informatics (ICHI)*, 443–444. IEEE.

Karim, F.; Majumdar, S.; and Darabi, H. 2019. Insights into lstm fully convolutional networks for time series classification. *IEEE Access* 7:67718–67725.

Kourentzes, N.; Barrow, D. K.; and Crone, S. F. 2014. Neural network ensemble operators for time series forecasting. *Expert Systems with Applications* 41(9):4235–4244.

Krishnan, P. T.; Joseph Raj, A. N.; and Rajangam, V. 2021. Emotion classification from speech signal based on empirical mode decomposition and non-linear features: Speech emotion recognition. *Complex & Intelligent Systems* 7:1919–1934.

Liu, C.; Ta, C. N.; Rogers, J. R.; Li, Z.; Lee, J.; Butler, A. M.; Shang, N.; Kury, F. S. P.; Wang, L.; Shen, F.; et al. 2019. Ensembles of natural language processing systems for portable phenotyping solutions. *Journal of biomedical informatics* 100:103318.

Liu, N.; Li, X.; Qi, E.; Xu, M.; Li, L.; and Gao, B. 2020. A novel ensemble learning paradigm for medical diagnosis with imbalanced data. *IEEE Access* 8:171263–171280.

Lotfidereshgi, R., and Gournay, P. 2017. Biologically inspired speech emotion recognition. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 5135–5139. IEEE.

Martis, R. J.; Acharya, U. R.; Lim, C. M.; Mandana, K.; Ray, A. K.; and Chakraborty, C. 2013a. Application of higher order cumulant features for cardiac health diagnosis using ecg signals. *International journal of neural systems* 23(04):1350014.

Martis, R. J.; Acharya, U. R.; Mandana, K.; Ray, A. K.; and Chakraborty, C. 2013b. Cardiac decision making using higher order spectra. *Biomedical Signal Processing and Control* 8(2):193–203.

Moody, G. B., and Mark, R. G. 2001. The impact of the mit-bih arrhythmia database. *IEEE engineering in medicine and biology magazine* 20(3):45–50.

Mosavi, A.; Sajedi Hosseini, F.; Choubin, B.; Goodarzi, M.; Dineva, A. A.; and Rafiei Sardooi, E. 2021. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. *Water Resources Management* 35:23–37.

Nabi, Z., and Nabi, Z. 2016. Machine learning at scale. *Pro Spark Streaming: The Zen of Real-Time Analytics Using Apache Spark* 177–198.

Nanni, L.; Lumini, A.; Dominio, F.; Donadeo, M.; and Zanuttigh, P. 2014. Ensemble to improve gesture recognition. *International Journal of Automated Identification Technology*.

Ortega, L. A.; Cabañas, R.; and Masegosa, A. 2022. Diversity and generalization in neural network ensembles. In *International Conference on Artificial Intelligence and Statistics*, 11720–11743. PMLR.

Osowski, S., and Linh, T. H. 2001. Ecg beat recognition using fuzzy hybrid neural network. *IEEE Transactions on Biomedical Engineering* 48(11):1265–1271.

Pan, M.; Zhou, H.; Cao, J.; Liu, Y.; Hao, J.; Li, S.; and Chen, C.-H. 2020. Water level prediction model based on gru and cnn. *Ieee Access* 8:60090–60100.

Parry, J.; Palaz, D.; Clarke, G.; Lecomte, P.; Mead, R.; Berger, M.; and Hofer, G. 2019. Analysis of Deep Learning Architectures for Cross-Corpus Speech Emotion Recognition. In *Proc. Interspeech 2019*, 1656–1660.

Ponti Jr, M. P. 2011. Combining classifiers: from the creation of ensembles to the decision fusion. In *2011 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials*, 1–10. IEEE.

Priyadharshini, M.; Banu, A. F.; Sharma, B.; Chowdhury, S.; Rabie, K.; and Shongwe, T. 2023. Hybrid multi-label classification model for medical applications based on adaptive synthetic data and ensemble learning. *Sensors* 23(15):6836.

Sagi, O., and Rokach, L. 2018. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8(4):e1249.

Sajjad, M.; Khan, Z. A.; Ullah, A.; Hussain, T.; Ullah, W.; Lee, M. Y.; and Baik, S. W. 2020. A novel cnn-gru-based hybrid approach for short-term residential load forecasting. *Ieee Access* 8:143759–143768.

Sangamnerkar, S.; Srinivasan, R.; Christhuraj, M.; and Sukumaran, R. 2020. An ensemble technique to detect fabricated news article using machine learning and natural language processing techniques. In *2020 International Conference for Emerging Technology (INCET)*, 1–7. IEEE.

Seijo-Pardo, B.; Porto-Díaz, I.; Bolón-Canedo, V.; and Alonso-Betanzos, A. 2017. Ensemble feature selection: homogeneous and heterogeneous approaches. *Knowledge-Based Systems* 118:124–139.

Shen, G.; Tan, Q.; Zhang, H.; Zeng, P.; and Xu, J. 2018. Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia computer science* 131:895–903.

Sigillito, V. G.; Wing, S. P.; Hutton, L. V.; and Baker, K. B. 1989. Classification of radar returns from the ionosphere using neural networks. *Johns Hopkins APL Technical Digest* 10(3):262–266.

Sundarprasad, N. 2018. Speech emotion detection using machine learning techniques.

Teh, Y. W., and Hinton, G. E. 2000. Rate-coded restricted boltzmann machines for face recognition. *Advances in neural information processing systems* 13.

Tsogbaatar, E.; Bhuyan, M. H.; Taenaka, Y.; Fall, D.; Gonchigsumlaa, K.; Elmroth, E.; and Kadobayashi, Y. 2021. Del-iot: A deep ensemble learning approach to uncover anomalies in iot. *Internet of Things* 14:100391.

Vanerio, J., and Casas, P. 2017. Ensemble-learning approaches for network security and anomaly detection. In *Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks*, 1–6.

Venkataramanan, K., and Rajamohan, H. R. 2019. Emotion recognition from speech. *arXiv preprint arXiv:1912.10458*.

Wang, Y.; Han, J.; Zhang, T.; and Qing, D. 2021. Speech enhancement from fused features based on deep neural network and gated recurrent unit network. *EURASIP Journal on Advances in Signal Processing* 2021:1–19.

West, D.; Mangiameli, P.; Rampal, R.; and West, V. 2005. Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application. *European Journal of Operational Research* 162(2):532–551.

Xu, Y.; Cheng, J.; Wang, L.; Xia, H.; Liu, F.; and Tao, D. 2018. Ensemble one-dimensional convolution neural networks for skeleton-based action recognition. *IEEE Signal Processing Letters* 25(7):1044–1048.

Yiğit, E.; Özkaya, U.; Öztürk, Ş.; Singh, D.; and Gritli, H. 2021. Automatic detection of power quality disturbance using convolutional neural network structure with gated recurrent unit. *Mobile Information Systems* 2021:1–11.

Yu, X.; Zhang, Z.; Wu, L.; Pang, W.; Chen, H.; Yu, Z.; and Li, B. 2020. Deep ensemble learning for human action recognition in still images. *Complexity* 2020:1–23.

Zhang, H.; Cheng, M.; and Hsieh, C.-J. 2019. Enhancing certifiable robustness via a deep model ensemble. *arXiv preprint arXiv:1910.14655*.

Zhang, S.; Zhao, X.; and Lei, B. 2013. Speech emotion recognition using an enhanced kernel isomap for human-robot interaction. *International Journal of Advanced Robotic Systems* 10(2):114.

Zhao, J.; Mao, X.; and Chen, L. 2019. Speech emotion recognition using deep 1d & 2d cnn lstm networks. *Biomedical signal processing and control* 47:312–323.

Zhao, Z.; Mehrotra, K. G.; and Mohan, C. K. 2015. Ensemble algorithms for unsupervised anomaly detection. In *Current Approaches in Applied Artificial Intelligence: 28th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2015, Seoul, South Korea, June 10-12, 2015, Proceedings 28*, 514–525. Springer.

Zhou, G.-B.; Wu, J.; Zhang, C.-L.; and Zhou, Z.-H. 2016. Minimal gated unit for recurrent neural networks. *International Journal of Automation and Computing* 13(3):226–234.

Zhou, F.-y.; Jin, L.-p.; and Dong, J. 2017. Premature ventricular contraction detection combining deep neural networks and rules inference. *Artificial intelligence in medicine* 79:42–51.