UBIQUITOUS PHYSIOLOGICAL PREDICTION OF SUD PATIENTS' WELLNESS STATE USING MEMORY-BASED CONVOLUTIONAL MODELS

Omid Dehzangi, Paria Jeihouni, Jad Ramadan, Victor Finomore, Nasser M. Nasrabadi, Ali Rezai

Rockefeller Neuroscience Institute, Lane Department of Computer Science and Electrical Engineering, West Virginia University, USA,

{{*omid.dehzangi, jramadan, victor.finomore, ali.rezai*}*@hsc, pj0001@mix, nasser.nasrabadi@mail*}*.wvu.edu*

ABSTRACT

The prevalence of substance use disorder (SUD) and rates of overdose in the United States have reached epidemic levels. Despite availability of effective evidence-based treatments for SUD, the rates of treatment attrition remain elevated. We have designed a cloud-based continuous physiological sensing for longitudinal SUD patient monitoring. Using wearable sensors, we aim to evaluate the impact of changes in heart rate (HR) and heart rate variability (HRV) signals on SUD wellness development using long-term and ubiquitous monitoring and machine learning and collected data from 10 subjects over an extended period of time. We designed a signal processing recipe and employed several recurrent neural network (RNN)-based architectures to track the temporal and spectral behavior of HR and HRV signals to predict the patients' wellness state. In addition, we have designed an architecture that combines RNN architectures and Time Scattered convolutional neural networks (TS-CNNs), where CNNs objectify the underlying features in the temporal dimensions within the signals (TS). The goal is to quantitatively and qualitatively evaluate the contribution of TS-CNNs in ubiquitous wellness prediction. The experimental results demonstrate that the best architecture configuration achieves 90.21% accuracy in predicting the wellness state of the SUD patients.

Index Terms— Wellness prediction, Wearable physiological monitoring, Time-Frequency analysis, RNN, CNN

1. INTRODUCTION

In in recent years, substance abuse, and overdose have reached epidemic levels in the United States. psychoactive substances derived from the opium poppy, have been reported as a substance, from which, 27 million people suffered in 2017 [1] and 118 thousand people died as a consequence of opioid overdose [1]. Furthermore, studies show that depression is highly correlated with increased levels of opioid addiction [2]. The patient's state observation is conventionally done by setting clinical appointments to check the biological metrics in the patients. Though, there is a need for continuous health state monitoring of the SUD patients. In addition, predicting the humans' wellness state, who previously were in the habit of using drug is of paramount importance as they could be reached out by urgent clinical help prior to any potential relapse.

With the advent of wearable sensors and internet of things devices (IoTs), pervasive and continuous patient state observation has been facilitated considerably (also in ur previous works) [3, 4, 5, 6], since the health data can be recorded ubiquitously and potentially transferred through their integrated cloud back-ends [7, 8, 9]. HR and HRV continuous recordings can shed light on the performance of the subject cardiovascular system and autonomous nervous system, respectively. We aim to compare the impact of changes in HR (representing active state) and HRV (representing resting state) signals on SUD wellness development. Recurrent neural networks (RNNs) are employed to deal with temporal features in the data. Various versions of the RNNs such as the long short-term memory (LSTM) and the gated recurrent unit (GRU) have been developed in different applications such as precipitation forecasting [10], natural language processing [11], video segmentation [12], etc. The distinctive characteristic of the RNNs is a certain type of memory they possess. Except simple RNNs using naive memory architecture, LSTM and GRU are equipped with a more advanced mechanism called gates which can regulate the flow of information coming from the earlier time steps. Hence, recurrent modules can process time-series data by considering the memory through time.

Another family of architectures is called convolutional neural network (CNNs), which is capable of detecting complex spatial features with different levels of abstractions in the data in hand [15, 16, 17, 18, 19, 20, 21]. Generally, CNNs are merely able to extract the spatial features out of the data, regardless of the temporal properties. Therefore, the integration of CNN and RNN seems to culminate in better perceiving the data, where CNN does the task of spatial feature extraction and RNN preserves the temporal properties. Xu et al. in [22], have employed CNNs followed by GRUs to perform the classification task for stereo recordings, and the method could achieve the state-of-the-art equal error rate. Also, authors in [26] could detect the QRS complex in ECG signal by feeding it to a CNN-LSTM architecture. Alzheimer's disease diagnosis was done in [27], where convolutional part extracts the features, and its combination with RNN could improve the model performance.

In this study, we propose a methodology to analyze con-

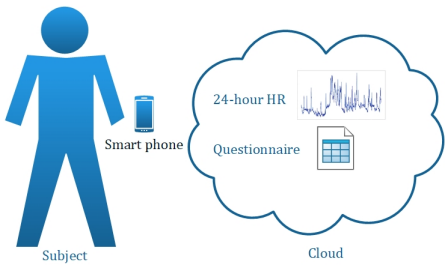

Fig. 1. Experimental setup designed for pervasive data collection.

tinuous 24-hour HR&HRV recordings for 150 days. Our aim is to investigate the relationship between physiological patterns, and the level of patient wellness by leveraging the capabilities of our proposed architecture of jointly-trained CNN-RNN structure. To the best of our knowledge, this is the first study quantifying the wellness of patients with respect to their physiological signals by using deep learning methods. If the subjetive measures show high correlation with the wearable objective indicators, that would be a crucial evidence of the usefulness of continuous monitoring of patients.

2. EXPERIMENTAL SETUP & DATA ACQUISITION

This study initially included 30 subjects with a history of substance abuse, whose ages were over 18. Each subject was given a Garmin vivosmart4 smartwatch, tracking their 24-hour HR&HRV for 150 days with a high sampling rate per day, which synchronized with their personal cell phone storing the data in a cloud using our in-house designed RNI health app. A graphic user interface was designed and installed on subjects' cell phones, on which they could report their depression levels through a daily questionnaire. The question's possible answers range from 1 to 5, implying the depression intensity in the subject. Experiments conducted in this work were approved by the Institutional Review Board (IRB) of the West Virginia University with the study ID: 1804090941. The outline of the experimental setup is depicted in Fig. 1.

2.1. Data Preparation & Feature Extraction

In this section, the policies selected for data preparation, pre-processing techniques, and feature extraction will be addressed in detail.

2.1.1. Data Framing

Let us assume that the physiological dataset collected for each patient is represented by $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_D\end{bmatrix}$, where D denotes the number of days, and $\mathbf{x}_i$ is a vector of the samples associated with the i^{th} day. Thus,

$$\mathbf{x}_i = \begin{bmatrix}{x_i}^1 & {x_i}^2 & {x_i}^3 & \cdots & {x_i}^N\end{bmatrix}^T, \tag{1}$$

where, $\mathbf{N}$ indicates the number of samples per day. Therefore, considering the sampling frequency to be equal to $f_s = 1/15^{Hz}$, the total number of samples per day would be $N = 24^{hours} \times 60^{mins} \times 60^{secs} \times 1/15^{Hz} = 5760^{samples}$. Furthermore, to prepare high-resolution data, the samples associated with each day are segmented with a window of 30 minutes of duration, and the overlap of $o.l. = 90\%$ between two consecutive windows. Hence, the number of windows per day would be:

$$M = \left\lceil \frac{N}{o.l.} \right\rceil. \tag{2}$$

Moreover, assuming the j^{th} window associated with the i^{th} day to be $\mathbf{u}_{i,j}$, and the length of each window to be l, then each frame of data follows the expression in Eq. (3).

$$\mathbf{u}_{i,j} = \begin{bmatrix} {x_i}^{(j-1)o.l.+1} \\ {x_i}^{(j-1)o.l.+2} \\ {x_i}^{(j-1)o.l.+3} \\ \vdots \\ {x_i}^{(j-1)o.l.+l} \end{bmatrix}. \tag{3}$$

Therefore, based on the Eq. (3), and 30 minutes duration considered for each window, the number of samples per window and the number of windows associated with each day would be $l = 120$ and $M = 480$, respectively. By segmenting the signal into smaller windows, in fact, we convert to small-time series chunks. Collecting real data in an unsupervised manner, we always need to enhance the quality of data prior to any further processing.

2.1.2. Data Pre-processing

Among HR&HRV samples collected by the Garmin vivosmart4 watches, there are usually a few missing values, also known as NAN values. Thus, to deal with the mentioned problems, the missing values are replaced by the median of the previous five samples. Finally, HR&HRV samples per day are framed into a consecutive set of 30-minute-long windows. However, it should be noted that the number of windows in a given day might differ from that of other days. Also, to assess the subjective wellness in terms of the depression level, a question is asked daily, for which the patients score their depression levels (questionnaire).

2.1.3. Feature Extraction

Following the data segmentation, feature extraction is carried out to achieve the highest possible efficiency out of the network. It stands to reason that the spectral features are highly representative of the signal behavior, as it maps the signal into the spectral domain. Fig. 2 illustrates temporal and spectral representations of the recorded signals. Since, the physiological signals are of one dimension, discrete Fourier transform (DFT) would be a candidate as the feature representation. Therefore, the spectral feature used in this study would

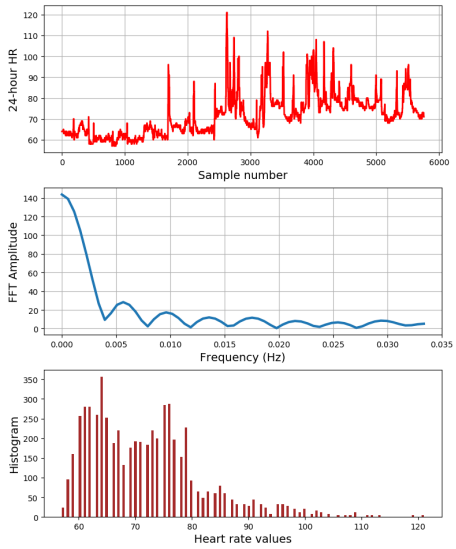

Fig. 2. From top to bottom: HR signal, spectrum, and histogram of signal values.

be the fast Fourier transform (FFT) of each time series frame, which is calculated by Eq. (4):

$$\mathbf{U}_{i,j}(k) = \sum_{n=0}^{l-1} \mathbf{u}_{i,j}(n) e^{-j\frac{2\pi}{l}nk}, k = 0, 1, ..., l-1 \qquad (4)$$

where, l denotes length of the j^{th} chunk for the i^{th} day. the signals are converted into time series windows and match with their corresponding depression levels.

3. METHODOLOGY

Two types of networks, employed in our proposed architecture for wellness state prediction, include the RNN and CNN. Using CNNs in the proposed method, we aim to conduct automatic feature extraction/detection in the physiological time series. Another structure is RNNs and its variants, whose mission is to preserve the temporal features of the physiological time series. We establish the methodology for the wellness prediction task based on a supervised learning paradigm, where the input is the physiological time series, its FFT representation, and the depression levels play the role of labels.

3.1. Time Scattered Convolutional Neural Networks (TSCNN)

In the first step, each chunk of the data is taken in by the convolutional layer. With that being said, each chunk needs to be fed to the network separately, and then distribute the data on continuous-time steps. In this study, each patch of data, i.e., group of windows fed to the network, contains the windows associated with only one day which corresponds to the label of that day. As depicted in Fig. 3, a signal chunk, covering 30 minutes of the data, undergoes a 1D-convolutional layer with 32 filters, receptive field of 96 samples, and with the rectified linear unit (ReLU) activation function. Thus, 32 feature maps with the size of 1×25 have resulted. After feature extraction, each feature map should be reduced in size to summarize the spatial features. Toward this end, each feature map is down-sampled using a global average pooling. This leads the network to have a single sample out of each feature map. Not only does the pooling layer reduce memory consumption, but also it helps concentrate more spatial information by averaging in the output, as well. Hence, as shown in Fig. 3 the final feature map extracted out of each chunk would be of a size 1×32. We have M windows per day. Therefore, the output of the convolutional layer associated with a given day would be of size $M\times1\times32$. As we need to feed each of the M output to the network separately, we call this layer time scattered convolutional neural network (TSCNN).

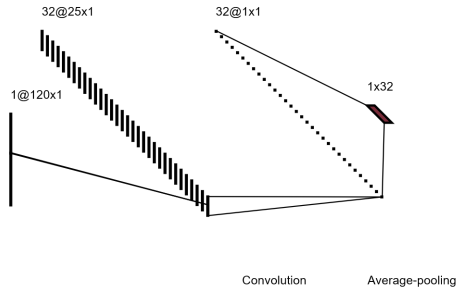

Fig. 3. Convolutional layer applied on a single chunk.

3.2. Recurrent Modules

Fig. 4 delineates a schematic structure for recurrent networks. As shown in Fig. 4, the RNN layers are stacked on top of each other to extract more temporal dependencies between different parts of the sequence to represent higher levels of abstractions (similar to the CNN networks). The major difference between the simple RNN with the LSTM and GRU versions is the gate mechanisms, embedded in the LSTM and GRU cells. These gates enable the network to be capable of dealing with only useful information, while forgetting the rest. Fig. 4 visually explains our entire proposed modeling architecture designed for the wellness state prediction. In Fig. 4, there is a batch normalization block (zero mean with unit variance). After passing the data to the stacked recurrent layers, the output of each unit in the last layer of the RNN directly goes to a fully-connected layer followed by a softmax to predict the wellness state of the subjects.

Table 1. The accuracy comparison among recurrent modules on temporal/spectral data, & in combination with TSCNN in(%).

Subject	Input Data	Sbj1 HR&HRV	Sbj2 HR&HRV	Sbj3 HR&HRV	Sbj4 HR&HRV	Sbj5 HR&HRV	Sbj6 HR&HRV	sbj7 HR&HRV	Sbj8 HR&HRV	Sbj9 HR&HRV	Sbj10 HR&HRV	AVE HR&HRV	AVE HR	AVE HRV
RNN	Raw	59.24	69.01	62.96	61.64	62.45	69.63	64.77	63.39	68.98	67.8	**64.98**	58.47	61.53
	FFT	64.15	68.37	60.37	62.73	77.71	73.12	67.48	70.62	75.27	65.86	**68.56**	62.35	65.48
LSTM	Raw	60.37	76.19	67.03	63.63	54.38	74.69	70.37	68.39	71.59	68.84	**67.55**	62.27	65.12
	FFT	67.92	72.06	77.61	69.69	75.43	76.43	65.72	73.05	78.5	73.83	73.02	70.67	**70.32**
GRU	Raw	64.15	71.42	77.19	81.23	78.78	75.71	74.49	76.47	80.46	67.03	**74.69**	68.84	72.34
	FFT	66.04	73.50	75.84	82.45	77.34	83.56	75.92	79.22	83.96	73.51	**79.09**	75.83	77.47
TS-CNN+RNN	Raw	64.28	65.08	31.57	58.82	56.61	73.87	71.82	65.7	69.04	64.06	**62.03**	59.16	60.59
	FFT	64.29	69.84	50.87	63.33	73.39	75.07	74.21	63.95	74.38	68.81	**67.81**	62.22	66.05
TS-CNN+LSTM	Raw	82.43	77.77	70.17	79.41	80.00	86.09	75.37	73.67	80.82	78.08	78.38	75.31	**78.69**
	FFT	72.85	77.77	80.70	79.17	75.96	87.36	77.98	76.85	80.17	81.34	**78.88**	71.98	73.14
TS-CNN+GRU	Raw	86.79	86.71	87.77	94.12	84.84	94.29	91.07	89.61	94.69	93.27	**90.21**	87.45	89.37
	FFT	69.81	61.11	84.21	85.29	81.83	81.67	75.52	83.46	88.25	89.01	80.01	77.69	**80.34**

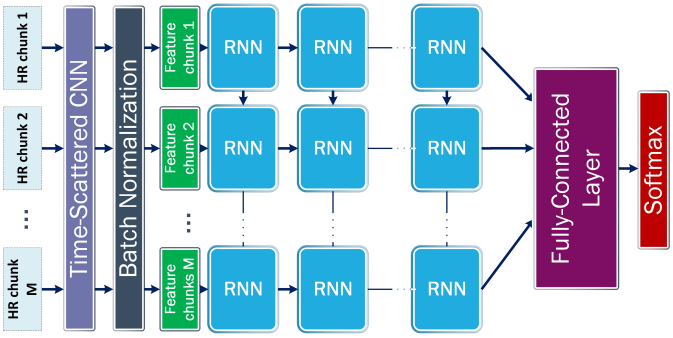

Fig. 4. Entire architecture for wellness prediction.

4. EXPERIMENTAL RESULTS

We employed massive data recorded from each subject individual and trained and evaluated the comparative models per each subject separately. The parameters set in the network are as follows: Learning rate= 0.001, loss function= *categorical cross-entropy*, optimizer function= *Adam*, batch size= 20, number of epochs= 200 and the number of recurrent layers= 4. Furthermore, we have evaluated the comparative models based on 10-CV and in a time series fashion.

4.1. Recurrent Modules Performance Analysis

The LSTM cells leverage three gates, i.e., input, forget and output, whereas the GRU cells keep the dependencies by the reset and update gates. The validation accuracies for all of the architectures are shown in the top part of Table 1 (with no convolutional layer). Table 1 reports individual subject models and their overall average performance columns on HR&HRV followed by average model accuracies for the cases of using only HR or HRV data for model training. Table 1 demonstrates that in most cases, the results using HRV signal as input is superior than using HR and their combination leads to the best prediction results. In Table 1, the best-achieved accuracies belong to the GRU and then LSTM as expected to perform better than simple RNN units. Table 1 also reports the prediction results comparing the temporal representation (i.e., raw signal) vs. the spectral representation (i.e., FFT transformed). As shown in Table 1, the FFT representation leads to better results compared to temporal representation that might be due to excessive noise in the time domain.

4.2. The Effect of the convolutional Layer on the results

The lower section of Table 1 presents the performance results on different recurrent modules in the presence of the time scattered CNN transform module. The results demonstrate considerable improvements compared to the previous results, which shows the effectiveness of the joint TS-CNN+LSTM model. In spite of the fact that LSTM has a more complex structure, with a lower computational complexity, GRU could achieve higher performance, e.g., 94.12% for the subject 4. Table 1 also reports the comparison between the raw vs. FFT representations and in this case, the raw input representation produces superior results unlike in top section of Table 1. This observation can be due to using CNN that conducts the feature extraction on the raw temporal data frames more effectively than the FFT that is a lossy process in terms of the temporal order of data in each window. Thus, the best architecture is TS-CNN+GRU with the raw (temporal) signal as the input to the network (Table 1).

5. CONCLUSION

In this paper, we have shown that by designing a deep learning method, specifically, using memory-based RNN methods, we can detect the underlying patterns in the HR & HRV signals and predict the SUD patients' wellness state in a longitudinal fashion. We examined several architectures and empirically found that the designed CNN + GRU architecture conducted on the input time series for feature extraction/objectification (CNN) and time series prediction (GRU), achieved the highest averaged prediction accuracy of 90.21%.

6. REFERENCES

[1] W. H. Organization: Management of substance abuse. World Health Organization, Tech. Rep. (2018).

[2] M. D. Sullivan: Depression effects on long-term prescription opioid use, abuse, and addiction. The Clinical journal of pain, 34(9), 878-–884 (2018)

[3] A. Pantelopoulos and N. G. Bourbakis: A survey on wearable sensorbased systems for health monitoring and prognosis. Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(1), 1-–12, IEEE (2010)

[4] S. C. Mukhopadhyay: Wearable sensors for human activity monitoring: A review. IEEE Sensors Journal, 15(3), 1321–1330 (2015)

[5] Mojtaba Taherisadr, Priyanka Asnani, Omid Dehzangi, ECG-based driver inattention identification during naturalistic driving using Mel-frequency cepstrum 2-D transform and convolutional neural networks, Smart Health, V. 9–10, pp. 50-61, (2018).

[6] Omid Dehzangi, Vaishali Sahu, Vikas Rajendra, Mojtaba Taherisadr, GSR-based distracted driving identification using discrete continuous decomposition and wavelet packet transform, Smart Health, V. 14, pp. 100085, (2019).

[7] O. Dehzangi, P. Jeihouni, V. Finomore and A. Rezai, "Physiological Monitoring Of Front-Line Caregivers For Cv-19 Symptoms: Multi-Resolution Analysis Convolutional-Recurrent Networks," 2021 IEEE International Conference on Image Processing (ICIP), pp. 250-254, (2021).

[8] O. Dehzangi et al., "XGBoost to Interpret the Opioid Patients' State Based on Cognitive and Physiological Measures," 2020 25th International Conference on Pattern Recognition (ICPR), pp. 6391-6395, (2021).

[9] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. Kwak: The internet of things for health care: A comprehensive survey. IEEE Access,3, 678–708 (2015)

[10] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.- c. Woo: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems, pp. 802–810, (2015)

[11] T. Young, D. Hazarika, S. Poria, and E. Cambria: Recent trends in deep learning based natural language processing. IEEE Computational intelligence magazine, 13(3), 55–75 (2018)

[12] D. Nilsson and C. Sminchisescu, "Semantic video segmentation by gated recurrent flow propagation: In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6819–6828, (2018)

[13] A. Graves, N. Jaitly, A. Mohamed, "Hybrid speech recognition with deep bidirectional lstm: IEEE Automatic Speech Recognition and Understanding, pp. 273–278 , (2013)

[14] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta: Long short-term memory network for remaining useful life estimation. In: 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), pp. 88–95, (2017).

[15] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. A. Heng: Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE Journal of Biomedical and Health Informatics, 19(5), 1627–1636, (2015)

[16] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng: An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE Journal of Biomedical and Health Informatics, 21(1), 31–40, (2017)

[17] S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, J. Cheng: Automatic scoring of multiple semantic attributes with multi-task feature leverage: A study on pulmonary nodules in ct images. IEEE Transactions Medical Imaging, 36(3), 802–814 (2017)

[18] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang: 3d cnn based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural mri. IEEE Access, 5, 23626–23636 (2017)

[19] G. Wang, et. al: Interactive medical image segmentation using deep learning with imagespecific fine tuning. IEEE Transactions on Medical Imaging, 37(7), 1562–1573 (2018)

[20] W. Song, S. Li, J. Liu, H. Qin, B. Zhang, S. Zhang, and A. Hao: Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition. IEEE Journal of Biomedical and Health Informatics, 23(3), 1215–1224 (2019)

[21] J. X. Qiu, H. Yoon, P. A. Fearn, and G. D. Tourassi: Deep learning for automated extraction of primary sites from cancer pathology reports. IEEE Journal of Biomedical and Health Informatics, 22(1), 244–251 (2018)

[22] Y. Xu, Q. Kong, Q. Huang, W. Wang, and M. D. Plumbley: Convolutional gated recurrent neural network incorporating spatial features for audio tagging. In IJCNN, pp. 3461–3466. IEEE, (2017)

[23] J. Zhao, X. Mao, and L. Chen: Speech emotion recognition using deep 1d 2d cnn lstm networks. Biomedical Signal Processing and Control, 47, 312–323 (2019)

[24] A. Satt, S. Rozenberg, and R. Hoory: Efficient emotion recognition from speech using deep learning on spectrograms. In: INTERSPEECH, pp. 1089–1093. Sweden (2017)

[25] W. Lim, D. Jang, and T. Lee: Speech emotion recognition using convolutional and recurrent neural networks. In: 2016 APSIP). pp. 1–4. IEEE, Korea (2016)

[26] B. Yuen, X. Dong, and T. Lu: Inter-patient cnn-lstm for qrs complex detection in noisy ecg signals. IEEE Access,7, 169359– 169370 (2019)

[27] C. Feng, et. al: Deep learning framework for alzheimer's disease diagnosis via 3d-cnn and fsbi-lstm. IEEE Access, 7, 63605–63618 (2019)