Introduction

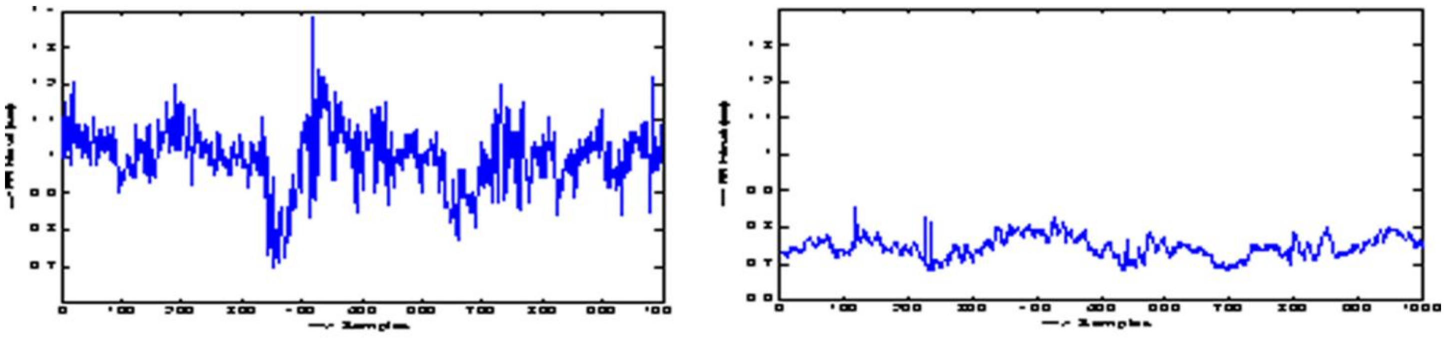
- Diabetes is a disease whereby blood sugar (glucose) is not metabolized in the body. This increases the glucose in the blood to alarmingly high levels. This is known by the name hyperglycemia. In this condition, body is unable to produce sufficient insulin. The other possibility is that body cannot respond to the produced insulin. Diabetes is incurable; it has to be controlled. A diabetic person can develop severe complications like nerve damage, heart attack, kidney failure and stroke. According to statistics in 2017, an estimated 8.8% of global population has diabetes. This is likely to increase to 9.9% by year 2045. In our present work, we analyse input HRV signals employing deep learning architectures of CNN, LSTM and its combinations. We achieve a high accuracy value of 95.7% employing CNN 5-LSTM architecture with SVM using 5-fold cross-validation. This work is the sequel to our published earlier work making use of deep learning techniques in diabetes detection with HRV as input data achieving an accuracy of 95.1%.

Summary of research works in diabetes detection with HRV data as input.

Authors	Methods	Accuracy obtained (in %)
Ref [4]	Nonlinear	86.0
Ref [5]	Higher order spectrum	90.5
Ref [6]	Higher order spectrum	79.93
Ref [7]	Nonlinear	90.0
Ref [8]	Discrete wavelet transform	92.02
Ref [9]	Empirical mode decomposition	95.63
Ref [3]	Deep learning (CNN-LSTM)	95.1
Proposed method	Deep learning (CNN-LSTM with SVM)	95.7

Description of dataset

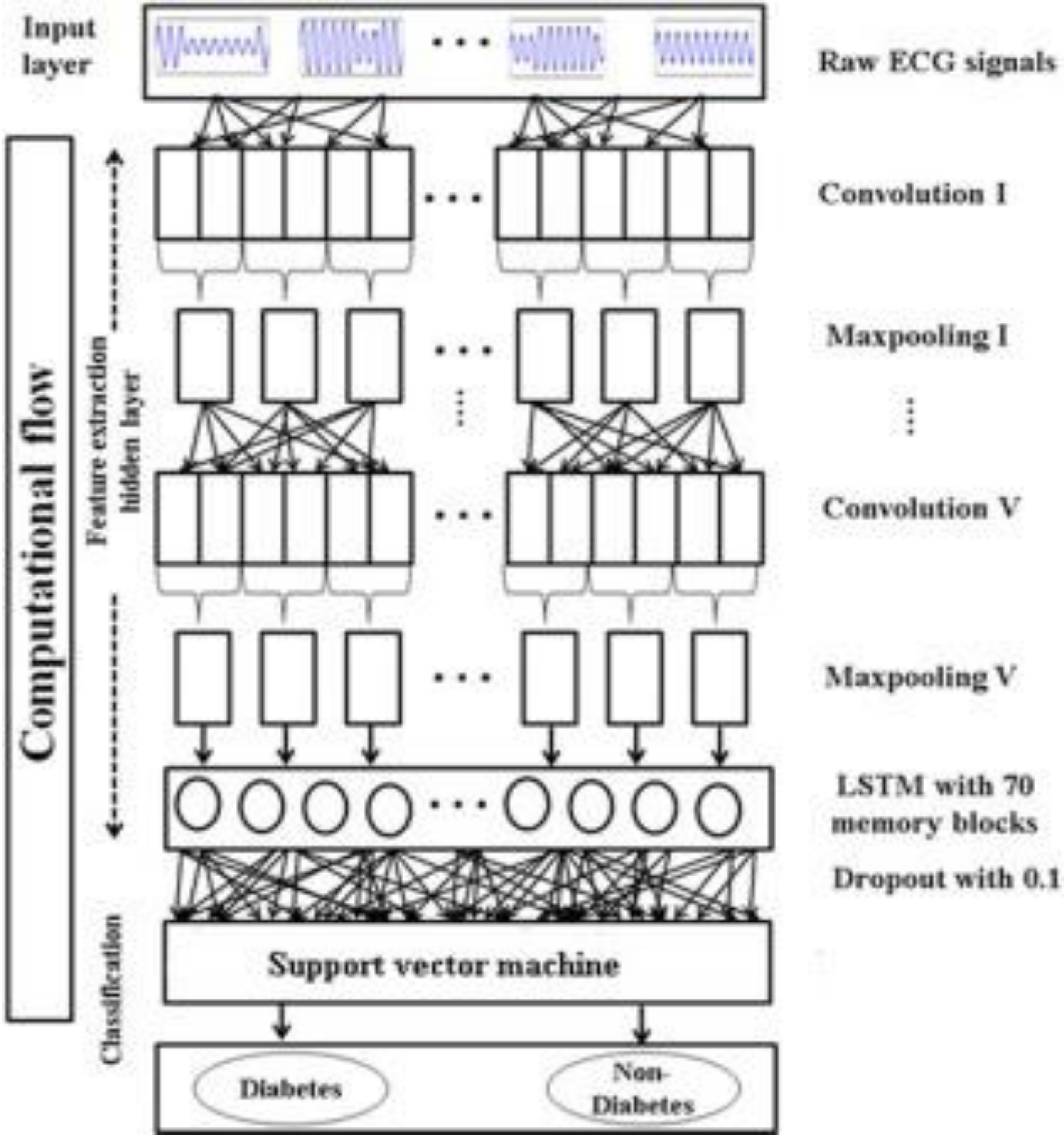
- The Electrocardiograms (ECG) of 20 people each from the diabetes and normal group were collected for 10 min with people lying down in a relaxed supine position. The heart rate time series data is derived from ECG signals using Pan and Tompkins algorithm. This real-time algorithm can effectively detect QRS complexes in an ECG signal based on its morphological features like slope, amplitude and width. It involves processes like digital bandpass filtering (to reduce false detections due to noise) and thresholding operations (to increase detection sensitivity). The ECG signal is sampled at 500 Hz. 71 datasets (same number for diabetic and normal group) each were extracted from the recorded data. Each dataset contains 1000 number of samples. The input data is passed to deep learning algorithms without any further pre-processing.



(A) (B)

Proposed architecture

- An overview of proposed architecture is shown . This is composed of 3 main sections. In input layer, the heart rate variability of raw ECG signal are given as input to deep learning architecture. This contains 5 CNN layers and each layer follows maxpooling. First two CNN layers contain 64 and 128 filters with filter length 3 and maxpooling with pooling length 2. Next two CNN layers contain 256 and 512 filters with filter length 3 and maxpooling with pooling length 4. A last CNN layer contains 1024 filters with filter length 3 and maxpooling with pooling length 6. This feature map is passed into LSTM layer. LSTM contains 70 memory blocks which learns the time domain features. This follows dropout 0.1. This randomly removes the neurons along with its connections. Finally, the features are passed into SVM for classification. The SVM used RBF kernel. This is defined as follows with samples s and s1.


Input layer
Raw ECG signals
Convolution I
Maxpooling I
Convolution V
Maxpooling V
LSTM with 70 memory blocks
Dropout with 0.1
Support vector machine
Diabetes
Non-Diabetes
Computational flow
Feature extraction hidden layer
Classification

Experiments and results

- All experiments are run on GPU enabled TensorFlow 10 with Keras 11 framework. In this work, we use the same configuration that we had used in our early paper . In this work, we extract features in deep learning network, comprised of CNN-LSTM architecture and pass into SVM for classification. LSTM has the capability to handle long-term dependencies in a data sequence. To decide the kernel function, we run two trail of experiment for SVM with linear and RBF kernel. SVM with RBF kernel performed better. These SVM model are implemented using Scikit-learn. The detailed 5-fold cross-validation accuracy is reported . In almost all the network structures, SVM has performed better in 5-fold cross-validation with accuracy which is comparable to the fully connected linear with nonlinear activation function for classification. Thus, we claim that the combination of SVM in penultimate layer for classification with deep learning layers for feature extraction can achieve the best performance.
-

Detailed results

Architecture	Accuracy obtained
CNN 1 with SVM	0.684
CNN 2 with SVM	0.755
CNN 3 with SVM	0.887
CNN 4 with SVM	0.913
CNN 5 with SVM	0.939
CNN 1-LSTM with SVM	0.743
CNN 2-LSTM with SVM	0.764
CNN 3-LSTM with SVM	0.853
CNN 4- LSTM with SVM	0.937
CNN 5-LSTM with SVM	**0.957**

Final Output

```
input_data = (5,166,72,19,175,25.8,0.587,51)

# changing the input_data to numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the array as we are predicting for one instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the input data
std_data = scaler.transform(input_data_reshaped)
print(std_data)

prediction = classifier.predict(std_data)
print(prediction)

if (prediction[0] == 0):
  print('The person is not diabetic')
else:
  print('The person is diabetic')
```

```
[[ 0.3429808   1.41167241  0.14964075 -0.09637905  0.82661621 -0.78595734
   0.34768723  1.51108316]]
[1]
The person is diabetic
```

Conclusion and future work

- Considerable part of human population is under the grip of diabetes which is incurable. If not managed well, diabetes can lead to health hazards. Hence, early detection of diabetes is extremely crucial. Nerve damages caused by diabetes, affect the working of the heart. In the proposed work, HRV data is analysed to diagnose diabetes using deep learning techniques. The maximum accuracy value of 95.7% was obtained for CNN 5-LSTM with SVM network. This is the highest value published for the automated diabetes detection with HRV as input data. Our non-invasive, flexible and reproducible system can serve as a reliable tool to clinicians to detect diabetes. Further improvement in accuracy can be obtained using a very large sized input dataset.
- The potential of deep learning is so tremendous that it can take a big stride in future to the so far challengingly difficult area of anomaly prediction from the anomaly detection if sufficiently large sized input data is available for research. The anomaly prediction can be tried from the input data which may not have anomaly by extracting dynamic characteristics from the input data. The predicted information can serve as a warning signal for the patient as well as the doctor to take sufficient control and precautionary measures.

References

- 1.
  Pfeifer M.A., Cook D., Brodsky J., Tice D., Reenan A., Swedine S., Halter J.B., Porte D.**Quantitative evaluation of cardiac parasympathetic activity in normal and diabetic man**

Diabetes, 31 (4) (1982), pp. 339-345

2.

- Kleiger R.E., Bigger J.T., Bosner M.S., Chung M.K., Cook J.R., Rolnitzky L.M., Stein man R., Fleiss J.L.**Stability over time of variables measuring heart rate variability in normal subjects**
- Am. J. Cardiol., 68 (6) (1991), pp. 626-630

3

- Swapna G., Kp S., Vinayakumar R.**Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals**
- Procedia Comput. Sci., 132 (2018), pp. 1253-1262

4.Acharya U.R., Faust O., Sree S.V., Ghista D.N., Dua S., Joseph P., Ahamed V.T. , Janarthanan N., Tamura T.**An integrated diabetic index using heart rate variability signal features for diagnosis of diabetes**

- Comput. Biomech. Biomed. Eng., 16 (2) (2013), pp. 222-234

- 5. Swapna G., Rajendra Acharya U., VinithaSree S., Suri J.S.**Automated detection of diabetes using higher order spectral features extracted from heart rate signals**
- Intell. Data Anal., 17 (2) (2013), pp. 309-326
- 6. Jian L.W., Lim T.-C.**Automated detection of diabetes by means of higher order spectral features obtained from heart rate signals**
- J. Med. Imaging Health Inf., 3 (3) (2013), pp. 440-447