Applied Project Final Report

Research Paper

Jonathan Paulick and Mitchell Mahnke

Milwaukee School of Engineering

Abstract

This paper will summarize our success in detecting congestive heart failure utilizing electrocardiogram (EKG) data. Our project built on the success of the paper “A New Deep Learning Method for Accurate Cardiac Heart Failure Prediction from R-R Interval Measurements,” in which the authors developed the T-LeNet architecture to detect cardiac heart failures. Our project explores the T-LeNet architecture and our model to high levels of success. As described in the results section, we achieved over 90% accuracy with both models.

Background

Our main inspiration for the final project was the first literature project. The paper we selected for our project is “A New Deep Learning Method for Accurate Cardiac Heart Failure Prediction from R-R Interval Measurements.” We chose this paper because it explores utilizing the R-R interval measurements from an EKG to determine how healthy an individual is. They looked at several deep learning techniques, showing how applying Deep Learning models is an iterative process. This focus aligns with our topic selection of estimation/prediction of illness using vitals. It should help highlight how the R-R interval, typically included when measuring vitals, can be used to predict some sort of issue, such as cardiac heart failure.

This research paper aims to find a technique that can reliably forecast the risk of cardiovascular illnesses. Their method of choice is deep learning, where they utilize algorithms such as Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) in order to make these predictions. They note a significant hurdle of a lack of medical data and that they are trying to beat out other algorithms that have already achieved a 93.1% accuracy rate. According to their results, we can also see that their difference in network structure, utilizing T-LeNet, allowed them to squeeze out an extra .9% accuracy rate. They note that these results are "experimental findings," which leads us to believe that their key contributions are the experimental findings, using T-LeNet and two accessible R-R interval databases to get an overall accuracy of 93.1%.

In our project, we also utilized PhysioNet for our datasets. PhysioNet is a resource established in 1999 by the National Institutes of Health (NIH) to conduct biomedical research and education. They offer free access to extensive collections of physiological and clinical data and open-source software. We specifically used the BIDMC Congestive Heart Failure Database and The MIT-BIH Normal Sinus Rhythm Database to create our condition and healthy dataset.

These datasets were not our original choice for datasets. We initially attempted to use the Congestive Heart Failure RR Interval Database and Normal Sinus Rhythm RR Interval Database. Unfortunately, we ran into a compatibility issue with the wfdb format. The files in these databases use the v1.3.9 version of wfdb, which is incompatible with current versions. There is no documentation for this older version, so we pivoted. The new dataset uses the newer digitized format. We needed to mention this as it is a significant learning experience about documenting versions when using non-human readable formats and knowing when to pivot. We found the version with the old datasets through trial and error.

Methods

We looked at different conditions for our project to be different from the original paper. Specifically, we looked at congestive heart failure as our condition. Our first goal was to adapt the model described in the original paper. The authors of the paper described their T-LeNet architecture by saying,

"The first convolutional layer will consist of 20 filters, followed by a max-pooling layer of length 2. The second convolutional layer will have 80 filters with a max-pooling layer of length 4. There will be 32 neurons used in the fully connected convolution layers to feed into the softmax classification layer."

This description is solid; however, the number of fully connected convolution layers is ambiguous. In our early effort to recreate this model, we ran into issues with output dimensions. We did not have a difficult time resolving this issue. As described in the result section, this original model did well on our new dataset.

After successfully implementing the model, we modified the architecture to improve our results. Our modifications include the following:

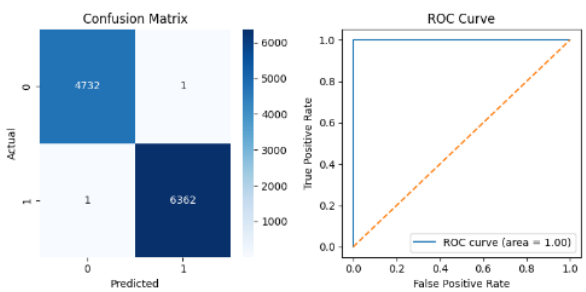
1. **Convolutional Layers:**
    - The original model starts with a convolutional layer with 20 filters, followed by a second convolutional layer with 80 filters.
    - The revised model employs three convolutional layers, with the first having 32 filters, the second having 64, and the third reverting to 32 filters.

2. **Dense Layers:**
    - In the original, there are multiple dense layers with 32 neurons each, directly following one another after convolutional and pooling layers.
    - The revised model has a single dense layer with 32 neurons before the output layer, which suggests a more streamlined architecture.
3. **Batch Normalization:**
    - The revised model introduces batch normalization layers after each convolutional layer, which are not present in the original. Batch normalization can help in stabilizing the learning process and reduce the number of epochs required to train the model.
4. **Pooling Layers:**
    - Both models use max pooling, but the revised model used them after each convolutional layer, whereas the original model has unspecified pooling layers.
5. **Flattening and Dropout:**
    - The revised model includes a flatten layer to convert the multi-dimensional input into a 1D array before passing it to the dense layers, which is a common practice in CNNs to prepare the data for fully connected layers.
    - A dropout layer is also added to the revised model, which is not present in the original. Dropout helps prevent overfitting by randomly setting input units to 0 at each update during training time.
6. **Parameters:**
    - The original model has 20,521 total parameters, all of which are trainable.
    - The revised model has significantly more parameters, with a total of 3,273,505 parameters, of which 3,273,249 are trainable and 256 are non-trainable.

Results

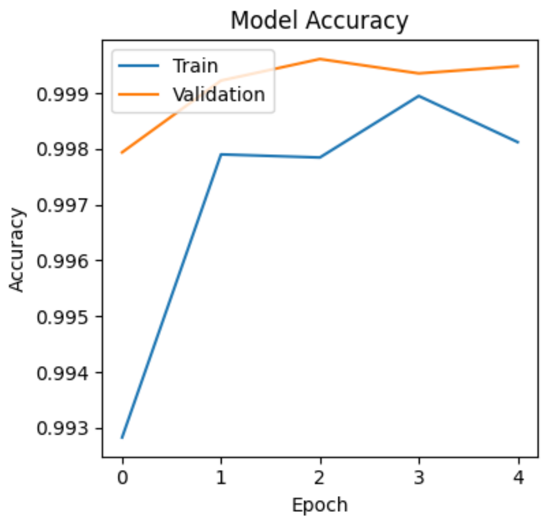
The implementation of our adapted T-LeNet architecture has yielded robust outcomes in the detection of congestive heart failure from EKG data. Our empirical analysis presents a comprehensive evaluation of model performance.

Enhancements to the T-LeNet model were instituted with the goal of elevating its diagnostic precision. These adjustments resulted in our modified model attaining an accuracy of 99%, as evidenced by the precision matrix that indicated a preponderance of true classifications, both positive and negative, with negligible false classifications. This underscores the model's capability to discern instances of congestive heart failure accurately.



	precision	recall	f1-score	support
NSR	0.99979	0.99979	0.99979	4733
CHF	0.99984	0.99984	0.99984	6363
accuracy			0.99982	11096
macro avg	0.99982	0.99982	0.99982	11096
weighted avg	0.99982	0.99982	0.99982	11096

The ROC curve achieved an AUC of 1.00, showing equality between positive and negative cases. Precision, recall, and F1-score for the NSR (normal sinus rhythm) and CHF (congestive heart failure) classes consistently exceeded 0.998, illustrating the model's acuteness in categorization and its equilibrium between sensitivity and specificity.



The model's accuracy trajectory over time was charted through epochal analysis, which indicated a consistent augmentation in predictive precision across both training and validation phases. The congruence of these trajectories suggests an absence of overfitting and a genuine learning progression.

The model's architecture has been refined to a total of 3,273,505 parameters, with trainable parameters and 256 non-trainable parameters. This refined parameter count, a reduction from the previous version, has not compromised the model's accuracy. On the contrary, the refined model demonstrates that a reduction in parameters can coincide with an improvement in performance metrics, a principle that resonates with the findings in the literature where increased accuracy was attained alongside a decrease in the number of parameters.

The balance between computational efficiency and accuracy is paramount, particularly for the prospective integration of such models into real-time diagnostic medical devices. Our model's enhanced parameter efficiency suggests a greater viability for deployment in resource-constrained environments, such as wearable health monitors.

Discussion

We successfully detected congestive heart failure with both the T-LeNet and our model. While we had high accuracy, our findings have little application due to how the medical field operates. The best case for our findings is that they inspire a similar project at a medical research facility.

From our experience in the medical field, the most significant hurdle is getting the project approved by a medical internal review board. This board reviews project requests to ensure ethics are at the forefront of the design. After a project is approved, much work is required to prove the model's effectiveness. If this model did find its way into a medical device, it would need to be qualified. This process ensures that any failure would not cause harm to the patient. As this model only does detection, it is unlikely that any harm could occur.

Our model outperformed the T-LeNet architecture in accuracy. One downside of our model compared with the T-LeNet is the total size. T-LeNet has 10,825 total parameters. Our model has 3,273,249 total parameters. This difference in total parameters does explain the improvements in accuracy. The increase in parameters would require a more powerful device for real-time predictions. Increasing the parameters only sometimes improves accuracy. In the original paper, the authors improved the accuracy and reduced the number of parameters.

Conclusion

In this project, we developed a model for successfully detecting congestive heart failure. Despite early data issues, we achieved 90% accuracy with the T-LeNet model and 99.98% with our model. This project taught us the importance of clearly documenting software versions if data uses a non-human readable format.

The next step with this project would be to try it with different conditions. PhysioNet has a large number of EKGs for different ailments. These datasets all use the same format, which would make plugging them into the notebook simple. We need to ensure that the wfdb files work with a modern version. Another project we could take on is reducing the number of parameters in our model. This change could allow us to port the model to wearable devices.

References

"BIDMC Congestive Heart Failure Database V1.0.0." *Physionet.org*, physionet.org/content/chfdb/1.0.0/.

"The MIT-BIH Normal Sinus Rhythm Database." *Archive.physionet.org*, archive.physionet.org/physiobank/database/nsrdb/.

"A New Deep Learning Method for Accurate Cardiac Heart Failure Prediction from RR Interval Measurements." *Ieeexplore.ieee.org*, 19 Nov. 2022, ieeexplore.ieee.org/document/10088409.

"Normal Sinus Rhythm RR Interval Database." *Archive.physionet.org*, archive.physionet.org/physiobank/database/nsr2db/. Accessed 11 Dec. 2023.

"Congestive Heart Failure RR Interval Database V1.0.0." *Physionet.org*, physionet.org/content/chf2db/1.0.0/. Accessed 11 Dec. 2023.