

SMARTWATCH BASED CONTINUOUS AUTHENTICATION WITH DEEP
LEARNING METHODS

by

Dilruba Köse - Furkan Aydar

Submitted to the Department of Computer
Engineering in partial fulfillment of
the requirements for the degree of
Bachelor of Science

Undergraduate Program in Computer Engineering
Boğaziçi University
Spring 2020

SMARTWATCH BASED CONTINUOUS AUTHENTICATION WITH DEEP
LEARNING METHODS

APPROVED BY:

Prof. Cem Ersoy
(Project Supervisor)

DATE OF APPROVAL: . .2020

ACKNOWLEDGEMENTS

We would like to express our gratitude to our project advisor, Cem Ersøy, and our primary supervisor, Deniz Ekiz, who guided us throughout this project. We would also like to thank our friends and family who supported us to collect data for the study.

ABSTRACT

SMARTWATCH BASED CONTINUOUS AUTHENTICATION WITH DEEP LEARNING METHODS

Increase in use of IoT devices and cloud services requires a new security solution in order to keep the valuable user info safe and private. This requirement arises from the fact that these technologies let users utilize their private info in fields such as digital banking, pervasive healthcare, media storage and transfer. In this paper, we offer a continuous implicit authentication solution for this purpose using Deep Learning models. Physiological signals obtained from wearable technological devices such as smartwatches provide an insight, as they are unique at certain time windows. EDA (electrodermal activity) and HRV (heart rate variability) are some of the good statistical candidates to be retrieved from smartwatches and to be used for authentication. We propose a solution to improve existing shallow machine learning solutions in this field. We will show that wearable devices can be used for continuous authentication for enhancing the security of the cloud, edge services, and IoT devices. We will also demonstrate that Deep Learning algorithms perform well as we got very low EER(Equal Error Rate) results.

ÖZET

DERİN ÖĞRENME İLE AKILLI SAAT TABANLI SÜREKLİ KİMLİK DOĞRULAMA

IoT cihazlarının ve bulut hizmetlerinin kullanımındaki artış, kullanıcıların kişisel ve önemli verilerini güvenli ve gizli tutmak için yeni güvenlik çözümleri gerekliliğini de beraberinde getirmiştir. Bu gereksinim, kullanıcıların özel bilgilerini, dijital bankacılık, online sağlık hizmetleri, medya depolama ve aktarma gibi alanlarda bu teknolojiler ile kullanmasından kaynaklanmaktadır. Bu çalışmada, Derin Öğrenme modellerini kullanarak sürekli kimlik doğrulama ile bu güvenlik gereksinimine bir çözüm sunuyoruz. Akıllı saatler gibi giyilebilir teknolojik cihazlardan elde edilen fizyolojik sinyaller, belirli zaman aralıklarında kişiye özel oldukları için kimlik doğrulamada kullanılabilir. EDA (elektriksel aktivite) ve HRV (kalp atış hızı değişkenliği) verileri akıllı saatlerden alınabilir ve kimlik doğrulamada kullanmak için iyi birer adaydır. Çalışmamızda, bu alanda daha önce yapılmış makine öğrenmesi çözümlerini geliştirmek ve giyilebilir cihazların sürekli kimlik doğrulamasında kullanılabileceğini göstermek amaçlanmıştır. Bu doğrultuda Derin Öğrenme modellerinin performanslarını kıyaslamak için EER (Eşit Hata Oranı) hesapladık ve aldığımız düşük hata oranları Derin Öğrenmenin hedeflenen performansı sağladığını gösteriyor.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ÖZET	v
LIST OF FIGURES	vii
LIST OF TABLES	viii
1. INTRODUCTION AND MOTIVATION	1
2. STATE OF THE ART	3
3. METHODS	5
3.1. Glossary	5
3.2. Data Collection and Research Process	5
3.3. Neural Network Models	7
3.3.1. RNN(LSTM)	7
3.3.2. CNN-LSTM	9
4. RESULTS	11
4.1. Performance of Deep Learning Models	11
4.2. Performance of the State-Of-The-Art and Comparison	13
5. CONCLUSION AND DISCUSSION	14
6. FUTURE WORK	15
REFERENCES	16

LIST OF FIGURES

Figure 2.1. Can A Smartband Be Used For Continuous Implicit Authentication in Real Life, overview	4
Figure 3.1. Early stopping in action. Network stops training before reaching maximum number of epochs (150 here), due to extinct improvement of validation loss.	6
Figure 3.2. An overview of the LSTM network	8
Figure 3.3. An overview of the CNN-LSTM network	10
Figure 4.1. Accuracy and EER results for CNN-LSTM and LSTM models . .	11
Figure 4.2. Accuracy and EER results for CNN-LSTM models with different window sizes	12
Figure 4.3. Accuracy and EER results for Proposed model and SOTA model .	13

LIST OF TABLES

Table 3.1. Architecture of our model. Output shape depends on specified window size (60 here).	8
Table 3.2. Architecture of our model. Output shape depends on specified window size (120 here).	9

1. INTRODUCTION AND MOTIVATION

Our research proposes a new method for continuous implicit authentication. Our aim is to provide a secure and reliable way for authenticating users via their wearable devices, utilizing various heart beat statistics obtained by also these devices. There are already work done in the field [1] by our colleagues in Bogazici University, in which they used Machine Learning methods. Yet we aim to employ a different approach. We are exploring Deep Learning methods for our work and try to achieve a higher success rate.

Security of the private user data stored in cloud services and IoT devices is a serious concern. It is an emerging need to keep critical and sensitive user info safe, since the leak of such valuable belonging would result in dramatic results or damage. As more and more people started to store and process their info through technological means, more opportunities for malevolent behaviour to abuse this precious pool of private data also has risen. This fact led the manufacturers of the mentioned means to take user concerns of the security issues more seriously. User authentication has become a crucial topic and one of the main parameters of the customer satisfaction, which introduced various applications and implementations of the authentication providers, such as password protection, face recognition, fingerprint recognition, voice recognition and more. Yet most of these implied technologies still have major drawbacks. Even strong kinds of the face recognition applications were exposed to be surmountable.

In most cases the problem is that the lack of authentication model's uniqueness, or the methods' imitability through some methods. This fact has motivated us to apply a more trustable and unique authentication method and merge it with existing technology. For us, utilizing human heart and body activity for security purposes would provide one of the most robust authentication method, according to the fact that these activities are unmatched, due to human body's nature. Also, emerging use of the IoT devices is one of the key motivators of our research. IoT devices such as smartwatches, which can track our body activity, like heart rate, are introduced to

our lives and being more popular day by day. Thus, it is an area that the continuous implicit authentication could be applied directly to, as these devices both need to be secured as they started to store valuable data and also it can track our body activity real-time.

2. STATE OF THE ART

Authentication with using biometric data has been increasingly used. There are different types of sensors used in research for recognition of individuals; fingerprint, voice, face, and electroencephalography. Each has its own advantages and drawbacks. Fingerprints and face-based authentication systems are available in most devices, but they can be easily deceived [2] [3]. Likewise, authentication with voice is available in most devices, but it has privacy issues. Other sensors like acceleration, gyroscope and, magnetometer, and electroencephalography used in the continuous authentication field are tested in laboratory environment and are not available in unobtrusive devices such as smartphones, smartwatches, and smart bands [4]. Also some sensors used in research are expensive and can be disturbing for the user in daily life usage.

Being also our research topic, it is recently realized that physiological signals are good candidates for continuous implicit authentication. The fact that signals such as variability in heart rate, blood volume pulse and electrodermal activity can be used for identification of unique users has shed light to some researches in the field. Before mentioning the state of the art of this topic, it is crucial to speak of some advantages of this method compared to other authentication mechanisms. First of all, physiological signals are difficult to deceive, intrinsically. It is almost guaranteed that a user's heart rate variability flow will not be replicated by any other user. Secondly, collected signals do not violate any privacy code. It is harder to obtain valuable private info from heart rates, compared to face or sound recognition, since the latter methods are more abusable and exploitable. Last but not least, sensors to record the signals needed are started to be shipped with most of the wearable smart products nowadays, so the idea is more applicable compared to others.

Currently, state of the art research [1] belongs to our supervisors Deniz Ekiz, Yekta Said Can, Cem Ersoy and our colleague Yağmur Dardağan, from Bogazici University. They have collected smartwatch data in same manner with our research. In their project, they have applied artifact removal to the obtained heart signals and then

fed these signals into Machine Learning algorithms. They have achieved 3.96% EER (Equal Error Rate) and concluded that HRV is a strong candidate for continuous unobtrusive implicit physiological authentication. A schematic overview of their work is below.

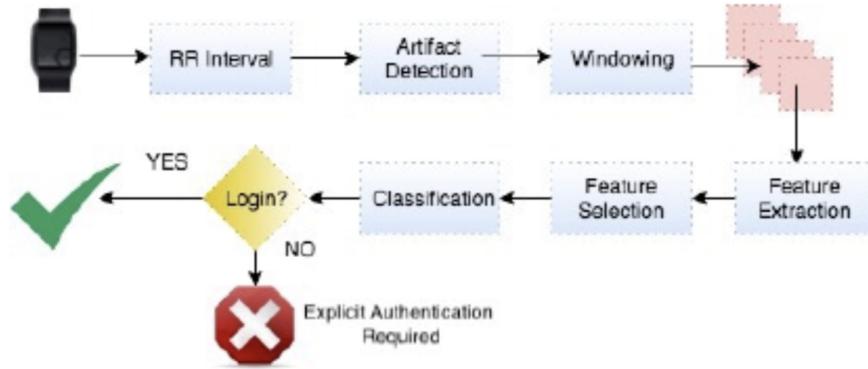


Figure 2.1. Can A Smartband Be Used For Continuous Implicit Authentication in Real Life, overview

Although their work is incandescent, it can still be improved. In our work, we replaced the shallow Machine Learning algorithms with Deep Learning algorithms and by doing so, we eliminated some steps above, like feature extraction and feature selection. We also tested and addressed the vitalism of the artifact cleaning process and some other operations such as refactoring the data accordingly to the BVP signals.

3. METHODS

3.1. Glossary

- Window size: Time period in which signal data of a single candidate are appended through. Main determinant of our input shape, fed into the neural networks. For every line of the input matrix, there are window size duration worth of data. This term also indicates the test period for candidate data. In our experiment we test if our model is able to identify a candidate in shorter window sizes (or time periods).
- Early Stopping: A method for halting the training process of neural network, under defined circumstances. In our research, we used early stopping to watch validation loss and stop training when validation loss stops decreasing for a while.
- EER: Equal Error Rate, an accuracy metric and verification system. Determined by the intersection point of FAR (False Acceptance Rate) and FRR (False Rejection Rate) of our model.
- Standardization (Standard Scaling): A method for transforming form of the dataset to have 0 mean and 1 standard deviation. This transformation is especially helpful to regularize datasets including positive and negative numeric values or data features with different units.

3.2. Data Collection and Research Process

There are two processes we conduct, which would account for the experiment phase.

(i) Training and Testing of our model

- We collect weekly data from volunteers by giving each participant a smart-watch and ask them to wear it 7 day in their daily life.
- We separate each participants weekly data to training (5-6 days) and test group (1-2 days).

- We split the data to window frames.
- We implement and set parameters of our neural network.
- For CNN-LSTM network, we applied Standard Scaling.
- We train our models to be specific to the users. In the training phase, if a data window frame belongs to the specific user, we set its output to 1, otherwise 0. Thus, our aim was to train a unique binary classifier for every participant.
- During training, we apply Early Stopping method.

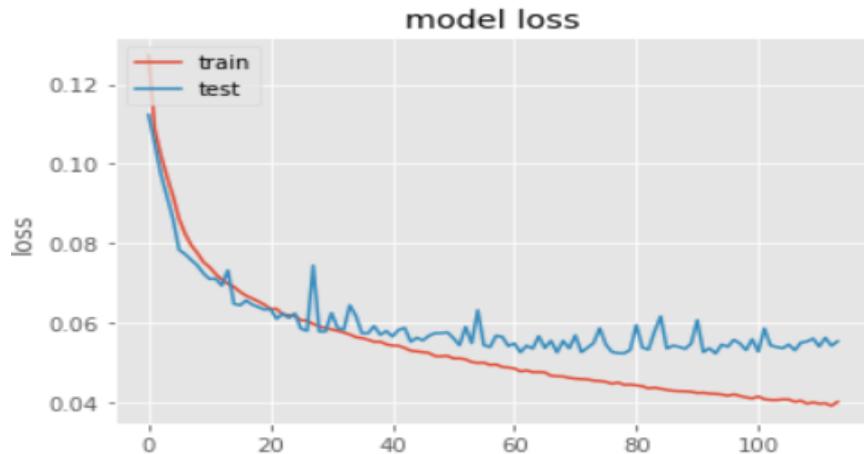


Figure 3.1. Early stopping in action. Network stops training before reaching maximum number of epochs (150 here), due to extinct improvement of validation loss.

- In the test phase, if a data window frame belongs to the specific user, we expect the test output to be 1, otherwise 0.
- Our models' successes are based on the expected test output and EER(Equal Error Rate). We expect to get higher average model success among all of our participants' models.

- During training, we split the data of candidates into 6 folds. For each fold, there were 5 training data point for every test data point. EER results of candidates are produced by averaging every fold's success.
 - During training, we split 20% of the training data to serve as validation set.
- (ii) Experimenting various neural network implementations
- Throughout this process, we try different neural networks (LSTM, CNN-LSTM, ConvLSTM) and different parameters (dropout rate, batch size, activation function etc.) and record the success rate. Here we demonstrate the models trained with best performing parameters.
 - As we split our data into window frames, we try different window lengths for this step.

Our research relies on the data which we have been collecting from the volunteers. We asked these participants to wear a smartwatch for a week and to upload their daily metrics which we will be utilizing. We also have been conducting surveys concerning the emotional and physical status of the attendants. The collected data is stored anonymously and the data collection procedure used in this study is approved by the Institutional Review Board for Research with Human Subjects of Bogazici University with the approval number 2018/16. Also, we asked the participants to sign a document that allowing us to use their collected data throughout our research. The document clarifies the topic, methods and goals of our work. Overall, we are sensitive about the consent of our participants and we have been trying to make them aware of the research's boundaries.

3.3. Neural Network Models

3.3.1. RNN(LSTM)

Our LSTM network consists of two LSTM Layers regularized with Dropouts and followed by two hidden Dense layers.

LSTM Network Architecture	
Layer Type	Output Shape
(LSTM)	(None, 60, 100)
(Dropout)	(None, 60, 100)
(LSTM)	(None, 100)
(Dropout)	(None, 100)
(Dense)	(None, 100)
(Dropout)	(None, 100)
(Dense)	(None, 1)

Table 3.1. Architecture of our model. Output shape depends on specified window size (60 here).

For LSTM model we aimed to have a deep model, so we used consecutive LSTM layers. We added Dropouts after each layer except the last Dense layer which is the output layer, in order to prevent over fitting. Since our data was unbalanced, it was mandatory to let the model forget some of what it learned and then let it learn again.

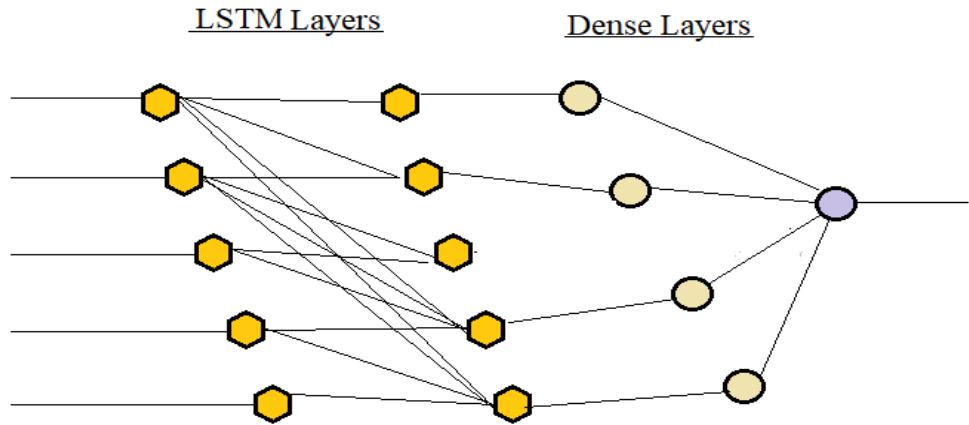


Figure 3.2. An overview of the LSTM network

Our LSTM model is a base for our next model, CNN-LSTM. It does not have any feature extraction phase, it directly learns from the input - window sized raw data- as shown in Figure 3.2.

3.3.2. CNN-LSTM

Our CNN-LSTM network consists of looping Convolutional Layers regularized with Dropouts and followed by a Bidirectional LSTM. Finally, we have two hidden Dense layers.

CNN-LSTM Network Architecture	
Layer Type	Output Shape
(TimeDistributed Conv1D)	(None, 1, 118, 256)
(TimeDistributed AveragePooling1D)	(None, 1, 39, 256)
(Dropout)	(None, 1, 39, 256)
(TimeDistributed Conv1D)	(None, 1, 37, 256)
(TimeDistributed AveragePooling1D)	(None, 1, 12, 256)
(Dropout)	(None, 1, 12, 256)
(TimeDistributed Conv1D)	(None, 1, 10, 256)
(TimeDistributed MaxPooling1D)	(None, 1, 3, 256)
(Dropout)	(None, 1, 3, 256)
(TimeDistributed Flatten)	(None, 1, 768)
(Bidirectional-LSTM)	(None, 384)
(Dropout)	(None, 384)
(Dense)	(None, 200)
(Dense)	(None, 1)

Table 3.2. Architecture of our model. Output shape depends on specified window size (120 here).

In our architecture -shown in Figure 3.3-, repeating convolutional layers serve to extract features, by treating consequent snapshots of biophysical signals as image inputs. In these layers we also analyzed effects of adding Max Pooling layers after every convolution operation. However, in first two layers of convolution, pooling by average resulted in deeper features extracted. As our data for every candidate is time distributed, averaging resulted in contribution of every data point included in the snapshots. For the final convolutional layer, we applied maximum pooling to retrieve the fluctuations of moving averages provided by two previous layers.

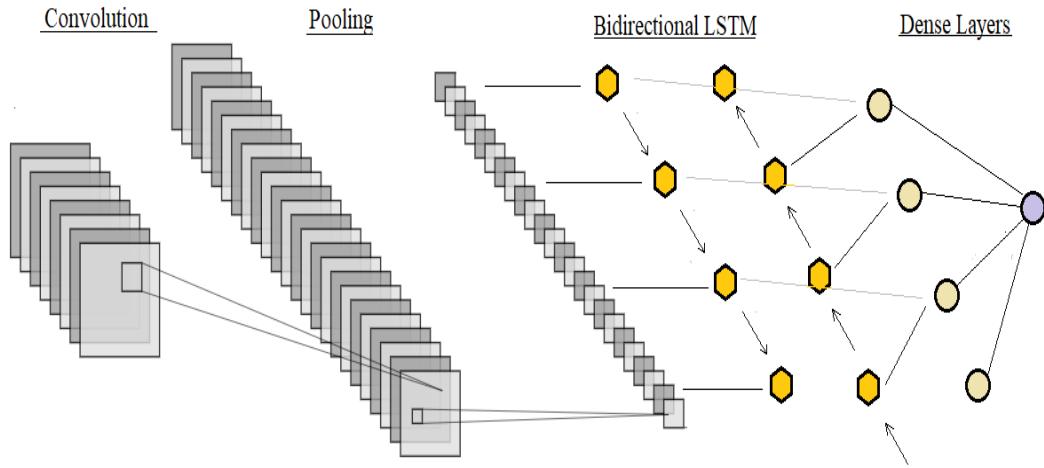


Figure 3.3. An overview of the CNN-LSTM network

Since our data set consists of various features calculated with constant intervals, we tried to leverage the power of sequential analysis of these features. For this purpose, we included LSTM's in order to incorporate the impact of certain timestamps on following and previous signals. As stated in [5], Regular LSTM has a shortcoming of just making use of previous context. For CNN-LSTM, we needed to analyze biophysical signals in two-direction for better analysis of feature flow. To illustrate, we were able to carry the effects of an arrhythmia observed at an arbitrary timestamp both to future and past sequences. Thus, by using Bidirectional LSTM's, our network was able to propagate features forward and backward.

Finally, output of LSTM layer is fed into a Dense layer. In order to make binary classification, we have another Dense layer with a single neuron.

To regularize the training and reduce overfitting, we applied Dropout to the outputs of Pooling layers and LSTM layer. This helps our network to randomly forget some features and weights in order to gain ability to enrich the discovered feature set.

4. RESULTS

We will discuss the results in two different sections. In the first one, we will present our Deep Learning models' results and will compare different models' success. In the second one, we will show the results of the state-of-the-art and then compare our results with these.

4.1. Performance of Deep Learning Models

We have applied many models and get the best results with LSTM and CNN-LSTM Deep Learning models. Here, we present 4 different models' results; LSTM with window size of 60 seconds, and CNN-LSTM with window sizes 30, 60 and 120 seconds. These results are calculated by testing each subjects' model against all subjects by taking samples from all subjects (Majority Subsampling). Accuracy and EER results are the averages of all subjects.

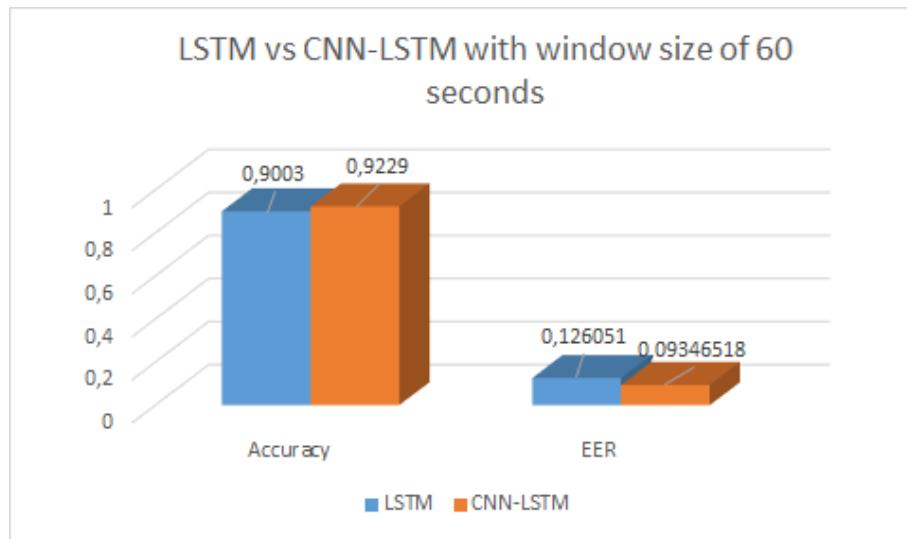


Figure 4.1. Accuracy and EER results for CNN-LSTM and LSTM models

The result of LSTM model and CNN-LSTM model with 60 window sizes is shown in Figure 4.1. LSTM with 60 window size has 90.03% accuracy while CNN-

LSTM with 60 window size has 92.29% accuracy. When we look the Equal Error Rates, we see that LSTM with 60 window size has 12.60% EER while CNN-LSTM with 60 window size has 9.35% EER. We can see that CNN-LSTM model has higher accuracy and lower EER than LSTM with same window size. It shows that CNN-LSTM model is more reliable since it has lower EER and higher accuracy.

In our training phase, we concluded that CNN-LSTM performed better than pure LSTM, so we tried different window sizes for CNN-LSTM model. In Figure 4.2., we show 3 different CNN-LSTM model with window sizes 30, 60, and 120.

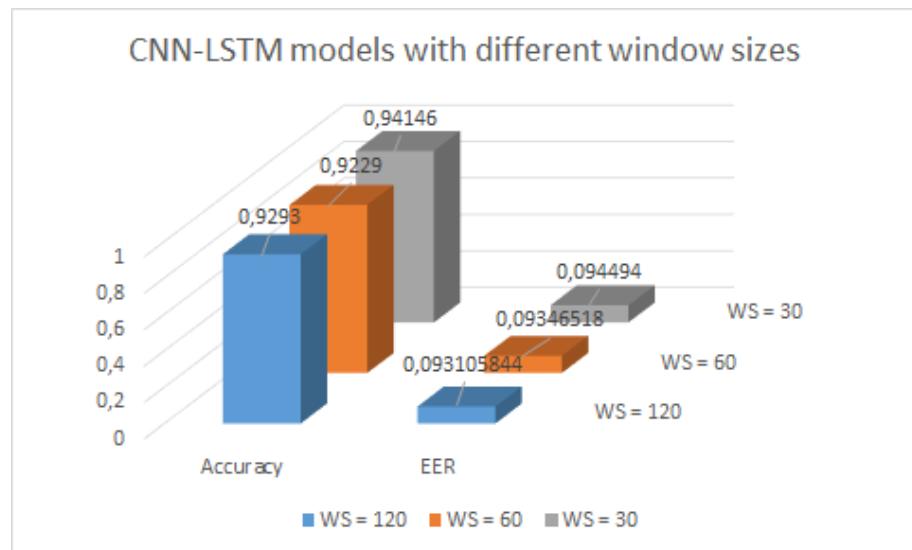


Figure 4.2. Accuracy and EER results for CNN-LSTM models with different window sizes

Here we see that CNN-LSTM has a lower EER as the window size increases. We cannot see the exact trend in accuracies. However, for us EER is more important because high EER means that modal has high False Acceptance and False Rejection rates which makes a model less reliable and more vulnerable to make faulty classifications. Also, we can conclude that our modal can predict a person's identity even within 30 seconds of window sizes, or namely test periods. We conclude that even within short test intervals, our model was able to identify candidates with high success

rate. Our network performed almost the same while defined test interval is reduced 4 times.

4.2. Performance of the State-Of-The-Art and Comparison

In state-of-the-art method, the best results achieved with Random Forest(RF) Machine Learning Algorithm. Here we present the comparison of best results of proposed algorithm and state-of-the-art(SOTA) algorithm in Figure 4.3. [1]. Proposed model has 92.93% accuracy while SOTA model has 98.48% accuracy. When we look the Equal Error Rates, we see that Proposed model has 9.31% EER while SOTA has 3.96% EER.

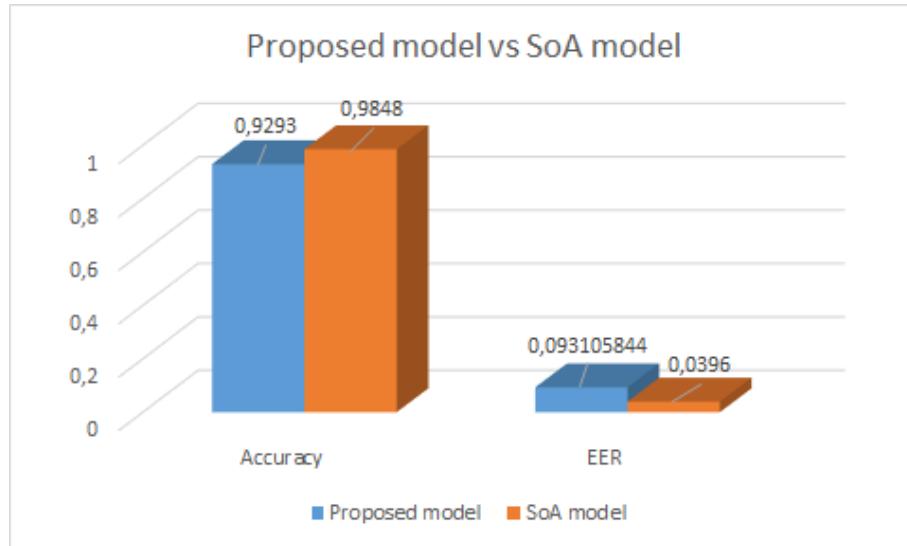


Figure 4.3. Accuracy and EER results for Proposed model and SOTA model

In their work [1], our colleagues analyzed the successes of different smart-watch models, and SOTA results in the Figure 4.3 reflects their best result obtained with Samsung Gear S. In our work, rather than analyzing according to the various watch models, we present an overall result of all the watch models. Even though our solution offers a higher EER, we have been able to show that continuous implicit authentication based on biophysical signals is possible even in shorter periods of sliding windows. We have also shown that our solution is able to present a reliable method with larger candidate pool.

5. CONCLUSION AND DISCUSSION

In this research, it is aimed to make authentication possible with physiological data collected from people's daily life. For this purpose, collected data is used in different Deep Learning models with different window sizes. We presented two different Deep Learning model; LSTM and CNN-LSTM. Even though we have achieved great sucess with some of the candidates, in average we are not in the position we wanted. We are currently facing varying EER rates for specific users' models, in rather wide range. This variety is probably a direct result of the lack of clear signal, since some candidates has shown extremely accurate authentication results, while some other's models have resulted in less accuracy. This outcome may result from the fact that we only used one kind of smartwatch and subjects used this watch in different time periods. As a result, we can say that CNN-LSTM with 120 window size can be a reliable continuous implicit authentication model since our model has relatively a high average accuracy, and low EER. Our models were able to maintain an acceptable success rate for classification, while the size of the dataset we work on exceed twice of the size compared to the [1].

6. FUTURE WORK

Currently, we have been successful applying specific deep learning approaches. We trained a deep sequential neural network and used LSTM and CNN-LSTM. On our collected data, our model was successful to determine if a data window belongs to a certain user to be tested. We think the larger our dataset gets, the more successful our project will be, since our model would be able to recognize more biometric data patterns.

If we would accomplish to reach our goal, a reliable method with high success rate would be introduced to authenticate users securely. The next step would be to apply our process to the different smartwatch brands and evaluate their success accordingly to their measuring powers. Currently, we are working with E4 smartwatches. Evaluating various competitors may result in fluctuation at the success of our model, as we currently are working with a brand that gives high precision results due to its focus on medical purposes.

Final destination would be deployment of our models to the wearable tech products used widespread, if our project is realized in a global context. Beyond its research value, we think that our project is highly applicable to the end products.

One possible improvement to our solution would be increasing the amount of the data collected from every candidate. (Collecting signals for two weeks rather than one). We also could utilize better methods for preprocessing the dataset to be even more suitable to our networks as input (Different interpolating techniques for cleaning unexpected or noisy signals). We think that our solution can be further improved to have product value and get integrated into various smartwatch models as an authentication API. Moreover, there is still more room to push the limits of smaller window sizes, in order to supply even faster authentication times.

REFERENCES

1. Deniz Ekiz, Y. C. D., Yekta Said Can and C. Ersoy, “Can A Smartband Be Used For Continuous Implicit Authentication in Real Life”, *IEEE Access PP(99):1-1*, Vol. VOLUME 4, 2016, March 2020.
2. Wencheng Yang, J. H. G. Z., Song Wang and C. Valli, “Security and Accuracy of Fingerprint-Based Biometrics: A Review”, *Symmetry 2019, 11, 141*, Vol. 11, 141, 2019.
3. Mary Grace Galterio, S. A. S. and T. Hayajneh, “A Review of Facial Biometrics Security for Smart Devices”, *Computers 2018, 7, 37;*, Vol. 7 ,37, 2018.
4. S. A. Elkader, M. B. and E. Lakshika, “Wearable sensors for recognizing individuals undertaking daily activities”, *ACM International Symposium on Wearable Computers - ISWC, ACM Press, 2018*, 2018.
5. Gridach, M., “Character-Aware Neural Networks for Arabic Named Entity Recognition for Social Medias”, , 2016.