APPLICATION OF A HYBRIDIZED LSTM-SVM IN THE DETECTION OF ARTIFACTS IN ELECTRODERMAL ACTIVITY SIGNALS FOR STRESS DETECTION

A Thesis Proposal

Presented to the Faculty of the College of Computer and Information Science

Polytechnic University of the Philippines

Sta. Mesa, Manila

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science

Benolirao, Johana

Calubayan, Christaline

Cueva, Larry Miguel

Quiray, Deseree

2024

| TABLE OF CONTENTS |  |
| --- | --- |
| **Table of Contents …………………………………………………………………...** | **1** |
| **Chapter 1 ……………………………………………………………………………..** | **2** |
| Introduction ……………………………………………………………………….. | 2 |
| Statement of the Problem ……………………………………………………….. | 10 |
| Hypothesis of the Study …………………………………………………………. | 10 |
| Theoretical Framework ………………………………………………………….. | 11 |
| Conceptual Framework ………………………………………………………….. | 13 |
| Scope and Delimitations of the Study …………………………………………. | 14 |
| Significance of the Study ………………………………………………………... | 15 |
| Definition of Terms……………………………………………………………….. | 15 |
| **Chapter 2 ……………………………………………………………………………..** | **19** |
| Related Literature ………………………………………………………………... | 19 |
| Stress Detection using Electrodermal Activity……………………………… | 19 |
| Electrodermal Activity Artifacts Detection using Machine Learning  Techniques……………………………………………………………………………. | 29 |
| Limitations of Machine Learning and Deep Learning Techniques in EDA Artifact Detection……………………………………………………………….. | 32 |
| Advantages of LSTM………………………………………………………….. | 38 |
| Hybridizing LSTM and SVM………………………………………………….. | 43 |
| Synthesis of the Study …………………………………………………………... | 55 |
| **Chapter 3 ……………………………………………………………………………..** | **57** |
| Research Design ………………………………………………………………… | 57 |
| Sources of Data ………………………………………………………………….. | 58 |
| Research Instrument ……………………………………………………………. | 58 |
| Procedures ……………………………………………………………………….. | 59 |
| System Architecture ……………………………………………………………... | 66 |
| Ethical Considerations …………………………………………………………... | 66 |
| Data Analysis (Procedure and Treatment) ……………………………………. | 66 |
| **References …………………………………………………………………………...** | **70** |
| **Appendix………………………………………………………………………………** | 84 |

# 

# 

# CHAPTER 1

**THE PROBLEM AND ITS SETTING**

## Introduction

Stress is defined as a condition that involves worry or mental strain generated by a challenging circumstance (World Health Organization, 2022). Stress can be found in different aspects or events. In the country of Jordan during the COVID-19 pandemic, the healthcare workers revealed that they are experiencing high level of stress and found out that 22.5% of participants had severe stress, 16.2% had extremely severe stress and 21.1% had a moderate stress in total of 60% of their healthcare workers experienced this level of stress (Alnazly et al., 2021).This high levels of stress experienced by healthcare workers during the pandemic can have negative impacts on their psychological well-being, job performance, and ability to provide quality patient care. With its negative connotations, numerous studies focused on detecting stress signals that may help in mitigating and managing stress to prevent and handle its effects better.

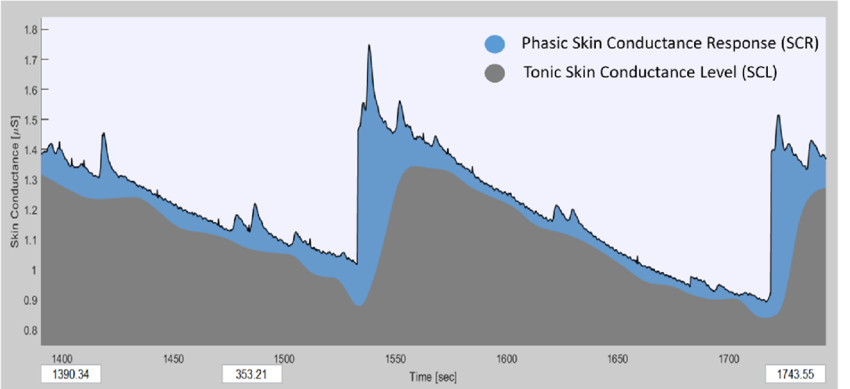
Coinciding with that, utilization of wearable sensors to detect psychological and physiological responses has been a trend over the course of years. In the investigation of Dzedzickis et al., (2020), the researchers conducted a review of sensors and models utilized for human emotion identification. In a publication by Semmlow (2004), many physiological processes produce energy that can be detected directly by a device generally known as a transducer which converts energy from one form to another. The energy that is converted by the input transducer may be generated by the physiological processes of the body itself. The measurement of electrical activity in the heart, muscle, or brain, provides other examples of direct measurement of physiological energy. For these measurements, the energy is already electrical and only needs to be converted from ionic to an electric current using an electrode. These include physiological processes that produce electrical energy, namely: electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG), electrooculography (EOG), and electrodermal activity (EDA).

Among the techniques used in measuring physiological responses is Galvanic Skin Response (GSR), also known as Electrodermal Activity (EDA) or more specifically, Skin Conductance Response (SCR) measured in microsiemens (μS).

Electrodermal Activity (EDA) is a low-cost and non-intrusive way of monitoring the emotional state of a subject, and a viable gateway to study the Sympathetic Nervous System (SNS), which is responsible for the so-called ﬁght-or-ﬂight responses happening at the unconscious level (Boucsein, 2012). EDA is also a frequently used modality in psychophysiology, because of its ability to obtain a distinct Electrodermal Response (EDR) in response to a stimulus. The possible uses of EDA are well documented; the most important examples include research on epilepsy, autism, stress and anxiety (Banganho et. al., 2022). More specifically because the phasic component or Skin Conductance Response (SCR) signal obtained from decomposing an EDA signal is closely associated with SNS activation when stimuli such as external stressors are present, numerous studies have emerged detailing how these signals can be exploited in order to detect stress within an individual. Therefore a brief overview of these studies detailing how indeed stress can be detected using only these physiological signals, will be followed in subsequent sections.

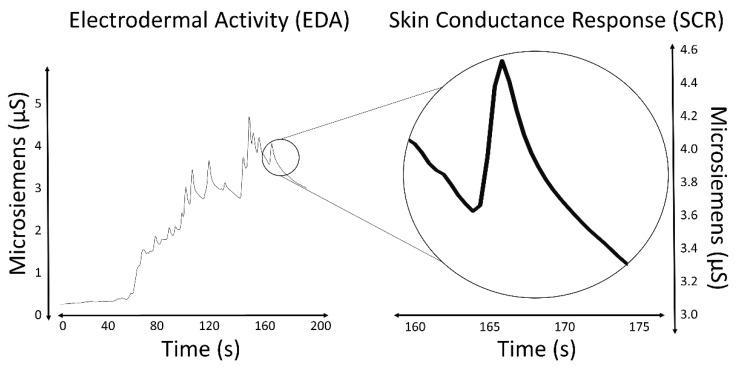
Emotional fluctuations cause sweating, which is most evident on the palms, fingers, and soles. This process alters the quantity of salt in the skin and changes its electrical resistance. This sensor provides less information about the emotional state compared to EEG and ECG, but it has a few advantages: it requires fewer measuring electrodes, which allows for easier use of wearable devices and definition of emotional states in ambulatory settings; it provides fewer raw data, which allows for faster analysis of obtained data; and it requires less computational power, and compared to other measurements, equipment needed for the measurement is much simple and cheaper.

Changes in electrodermal activity (EDA) can be either short-term (phasic) or longer-lasting (tonic). Phasic changes, known as skin conductance responses (SCR), are quick shifts in electrical resistance occurring within seconds. In contrast, tonic changes, measured as skin conductance level (SCL), represent EDA over a longer period (Boucsein et al., 2012).



*Figure 1. Components of EDA: Phasic (Skin Conductance Response) and Tonic (Skin Conductance Level)*

On the other hand, Jose (2022) has stated that within the domain of statistical computing, time series analysis is a statistical technique that deals with trend analysis and time series data. Time series analysis made its way into medicine when the ﬁrst practical electrocardio-grams (ECGs), which can diagnose cardiac conditions by recording the electrical signals passing through the heart, were invented in 1901. Time series analysis is a speciﬁc way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording

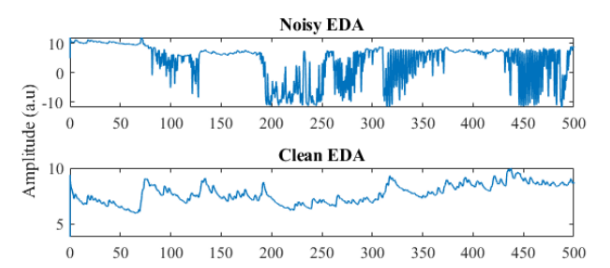


*Figure 2.Skin Conductance Response relative to time*

the data points intermittently or randomly. Given that signals recorded from ECGs, EEGs, EMGs, EOGs, and GSRs/EDAs share similar features in that they are sequences of data points gathered over time, GSR or EDA data may be classified as time series specific data. However, similar to other physiological signals, EDA signals face challenges. With the recent trend of wearable technologies for psychophysiological monitoring in ambulatory settings, the data quality from these sensors can be compromised by "noise" or artifacts in long-term recordings. Artifacts, which are unwanted changes in recorded biosignals not originating from the intended source (Boucsein, 2012), can result from unstable electrode contact, environmental factors like temperature and humidity, or movement. The referenced and noisy EDA channels are almost similar in shape, although different in amplitude (Hossain et al., 2022).

According to Boucsein (2012), detecting these artifacts typically requires visual inspection of the data, but this is unreliable in terms of large-scale EDA studies and long-term monitoring outside clinical settings. While signal processing techniques like low-pass filtering can be used to reduce the need for visual inspection, they risk altering the physiological response, potentially transforming the entire EDA trace and making artifacts appear as genuine responses.

If these noises persist in the signal when it is examined, they are susceptible to misunderstanding and distort the results; for example, they could be confused with a skin conductance response (SCR), a physiological reaction that may indicate increased stress (Taylor et. al., 2015). Recent studies have focused on developing models to automatically identify and remove artifacts (Gashi et al., 2020).

**

*Figure 3. Example of Noisy and Clean EDA signal from the study of Hossain et. al. (2022)*

In the study conducted by Llanes-Jurado et al. (2023) on automatic recognition and elimination of artifacts in electrodermal activity (EDA) signals using their EDABE dataset, they collected data from 43 participants in a stress-inducing VR study. The researchers developed and trained four models, two of which replicated traditional machine learning methods by Taylor et al. (2015) and Hossain et al. (2022).

Replicating the Taylor et al. (2015) method, they extracted 62 hand-crafted features, selected 40 via backward selection with a Support Vector Classifier, and used three classifiers: Logistic Regression, Support Vector Classifier, and Random Forest. They performed hyperparameter tuning and selected the model with the highest accuracy.

Additionally, they also had reimplemented the same methodology used in the recent paper by Hossain et al. (2022), where instead of 5s segments as it was in the latter paper, they used 0.5s segments of EDA signals. They engineered typical statistical features as with the paper by Taylor et al. (2015). The researchers included the optimized coefficients of an autoregressive model as features, excluding the bias or intercept coefficient. Finally, they used two time frequency transformation methods to extract time frequency features in order to capture non-stationary characteristics from the signals. A total of 50 features were engineered and extracted from the raw EDA signal data, then reduced to 40 using a Random Forest classifier as a feature selection method. The features before being fed as input for Support Vector Machine, Gradient Boosted Tree, Random Forest, and Logistic Regression classifiers were standard scaled and normalized using min-max. In order to select the best model for each classifier, the use of hyperparameter tuning has been repeated for each classifier together with the use of 5-fold cross validation to select the best model out of each classifier. The model that had the highest accuracy out of each classifier category was defined as the best model.

In addition to these, they proposed new models, which includes LSTM with a 1D-CNN and a 2D-CNN for analyzing signal spectrograms. The LSTM-1D CNN model recognized 72% of artifacts with 76% AUC (area under curve), 57% F1-Score, and 88% accuracy on the test set. Future work includes adding expert manual correction, developing movement protocols, and fine-tuning model architectures. This study reaffirmed that SVM, Gradient Boosted Tree, and Random Forest classifiers achieved the best accuracy on the validation set, consistent with previous findings.

In another study of Lee et al. (2020), artifact detection is a crucial aspect addressed in the research. The denoising method proposed in the study focuses on alleviating intrinsic respiration noise and extrinsic noise in Electrodermal Activity (EDA) signals collected by a wearable biosensor. The method involves detecting and attenuating irregular respiration-induced noise in EDA signals. Specifically, irregular respiration is identified and removed as respiration noise using a machine learning model that detects irregular respiration patterns from Photoplethysmography (PPG) signals collected simultaneously. This artifact detection process is essential for improving the accuracy of stress measurement by reducing noise interference in EDA signals, thus enhancing the reliability of stress metrics extracted from the data.

Despite these findings, traditional machine learning methods have limitations in healthcare signal processing. Sun et al. (2018) found that hand-designed EEG feature extraction methods resulted in poor analytical performance. They addressed this by using recurrent autoencoders for feature extraction. Hussein et al. (2018) further supported this by employing LSTM networks to capture high-level patterns in EEG signals. In their approach, a fully connected layer was used to extract robust, epileptic-relevant features, while a softmax layer provided predicted labels. This method maintained high detection performance, particularly in identifying artifacts like eye movements, muscle movements, and background noise. Additionally in a study by Jamshidzadeh et al. (2023) one of these limitations with regards to the use of traditional ML methods was specifically the inability of SVM models to extract significant features from data.

Moreover, while it has been established that convolutional neural networks achieve superior performance on high dimensional data such as images and used in different automated tasks, such as classification, detection, segmentation, data augmentation Szegedy et al. (2016) CNNs however according to Wang et al. (2020), Park & Yang (2019), Li et al. (2022) are still considered as a ‘Black box’ in terms of the underlying mechanism which makes it difficult to interpret the results and have confidence that they provide the optimal solution. A paper by Azam et al. (2023) also states that the ‘black box’ nature of CNNs is a key research interest currently where different research has attempted to provide different insight. There are several research questions related to the ‘black box’, such as the inner mechanism of CNN layers, feature interpretation and decision making schemes and as such CNNs are still poorly understood, not only by non-technical users but also by experts. This lack of knowledge according to Lange et al. (2018) may cause ambiguity and a hesitance in relying on the predictions of CNNs, especially in critical applications like the medical domain. And by opening up the ‘black box’ it can increase the confidence of users such as medical specialists in the results of neural networks as per studies conducted by [Ferdinand & Mercier, (n.d.](https://www.sciencedirect.com/science/article/pii/S2667305323000583?fbclid=IwZXh0bgNhZW0CMTAAAR237Nav-d71WwmVzTLNLOBqE1XXAkd6BSOqSv_C7nxX_ucXZ_Q7_5pZ-3o_aem_AQigPAAWF8bKS27hMiDqyvmg_qlmIqHGDdt3kI_JnosTqgs_apWmIkyg3SSq1gQ4LIR6basuDJs93-vtpaEE7NA7#bib0012)), [Brahimi et al. (2018](https://www.sciencedirect.com/science/article/pii/S2667305323000583?fbclid=IwZXh0bgNhZW0CMTAAAR237Nav-d71WwmVzTLNLOBqE1XXAkd6BSOqSv_C7nxX_ucXZ_Q7_5pZ-3o_aem_AQigPAAWF8bKS27hMiDqyvmg_qlmIqHGDdt3kI_JnosTqgs_apWmIkyg3SSq1gQ4LIR6basuDJs93-vtpaEE7NA7#bib0004)), and [Dependent et al. (2021](https://www.sciencedirect.com/science/article/pii/S2667305323000583?fbclid=IwZXh0bgNhZW0CMTAAAR237Nav-d71WwmVzTLNLOBqE1XXAkd6BSOqSv_C7nxX_ucXZ_Q7_5pZ-3o_aem_AQigPAAWF8bKS27hMiDqyvmg_qlmIqHGDdt3kI_JnosTqgs_apWmIkyg3SSq1gQ4LIR6basuDJs93-vtpaEE7NA7#bib0008)).

Finally, to address these challenges present in the aforementioned studies the researchers propose a hybridized LSTM-SVM model, where the LSTM aims to address the shortcomings of traditional ML methods of feature extraction from time series data, specifically those involving EDA signals, and wherein the integrated mechanism of traditional ML-based methods to the LSTM namely the Support Vector Machine (SVM) aims to address the shortcomings of the use of difficult to interpret architectures of CNNs.

## Statement of the Problem

The objective of this study will be to gauge the performance of our proposed LSTM-SVM model against other models that have already been established previously, and whether or not its performance will be significantly greater than that of the other models. In terms of gauging the performance the study asks the following questions:

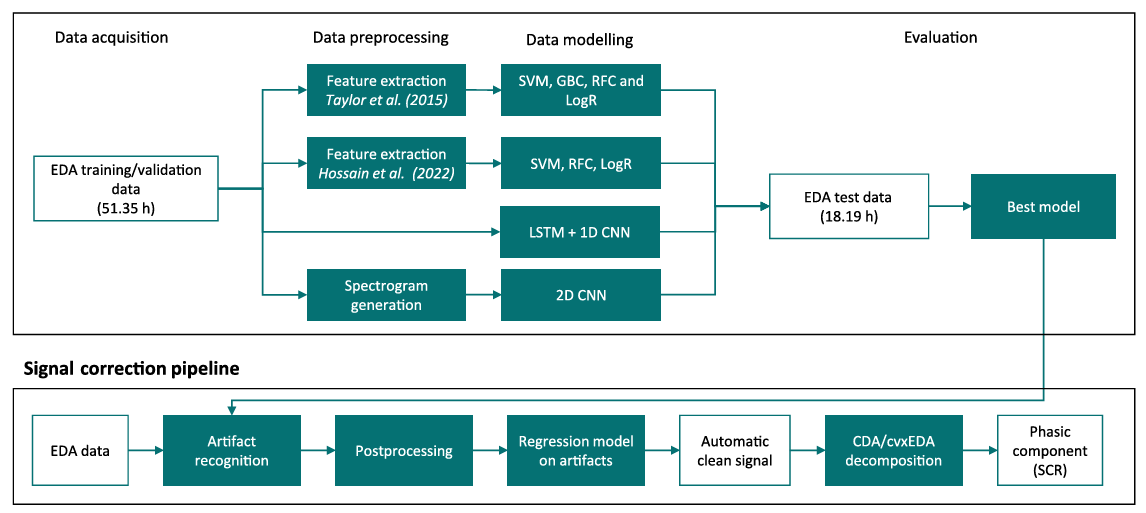
1. What would be the performance of the hybridized LSTM-SVM model compared to other existing models (SVM, GradBoost, LSTM-CNN) that automatically detect artifacts in EDA signals in terms of its:
2. AUC
3. Precision
4. Recall
5. F1-Score
6. Accuracy
7. What would be the optimal Gamma and C hyperparameters for the SVM mechanism of the model to achieve higher validation metrics as opposed to the other state of the art automatic artifact detection methods such as SVM, Random Forest, Gradient Boosted Tree, and LSTM-CNN?

**Null Hypothesis (H0)**

There is no significant difference between the performance of the hybridized LSTM-SVM model compared to the existing automatic detection models of SVM, GradBoost, and LSTM-CNN in terms of AUC, F1-score, accuracy, precision, recall.

## Theoretical Framework

The purpose of this study is to improve the detection of artifacts/noise in electrodermal activity (EDA) signals by utilizing advanced machine learning methodologies, specifically through a hybridized Long Short Term Memory (LSTM) infused with Support Vector Machine (SVM). This strategy will be contrasted with conventional machine learning approaches like Support Vector Machines, Linear Regression, Random Forests, which have been previously utilized in stress recognition research.



*Figure 4. Theoretical Framework*

1. **Machine Learning and Deep Learning Theories**

**1.1. LSTM & SVM:** Long Short Term Memory (LSTM) networks and Support Vector Machines (SVM) are both prominent techniques in the realm of machine learning, each offering distinct advantages in handling sequential and non-linear data, respectively (Cimino & Dell’Orletta, 2016). The hybridization of LSTM with SVM offers a promising approach to utilize the strengths of both models for improved artifact/noise detection in electrodermal activity (EDA) signals, particularly in the context of stress detection.

1. **Signal Processing Theory**

**2.1. Electrodermal Activity (EDA):** The measurement of Electrodermal Activity (EDA) involves assessing the skin's electrical conductance, which fluctuates in response to sweat gland activity and is modulated by the sympathetic nervous system. This method is frequently utilized in the identification of stress and in psychophysiological investigations. For a precise analysis of EDA signals, it is imperative to conduct efficient artifact removal procedures to ensure that the collected data accurately represents genuine physiological responses, rather than being distorted by noise or artifacts stemming from motion or environmental factors.

**2.2. Artifact Detection in EDA Signals:** The presence of artifacts within EDA signals can have a notable impact on the precision of stress detection models. Traditional techniques for identifying artifacts typically entail manual scrutiny or the application of heuristic algorithms, which can be labor-intensive and may yield suboptimal results (Gashi et al., 2020).

1. **Stress Detection Framework**

**3.1. Machine Learning in Stress Detection**: Conventional machine learning techniques such as Support Vector Machine, Linear Regression, Random Forests, and Naive Bayes have been extensively utilized for stress identification utilizing physiological signals. These approaches usually necessitate thorough feature manipulation and may not entirely capture the temporal dynamics of EDA signals. In contrast, Sequence-to-Sequence models can acquire knowledge directly from unprocessed data, potentially resulting in enhanced performance in identifying and rectifying artifacts.

**3.2. Comparative Analysis:** This analysis will evaluate the efficiency of the proposed LTSM-SVM model in comparison to the already existing methods of Taylor et al. (2015), Hossain et al. (2022), and Llanes-Jurado et al. (2023). Essential criteria for evaluation will encompass AUC, F1-Score, precision, recall, and accuracy of baseline classifiers. The claim argues that the hybridized LSTM-SVM model will excel over conventional techniques in anomaly detection within EDA signals by effectively utilizing contextual information from both preceding and subsequent data points.

Through analyzing these models with conventional artificial intelligence methodologies, this investigation seeks to enrich the domain of tension identification by enhancing the precision and dependability of EDA signal analysis. This framework will guide the research design, data collection, analysis, and interpretation, ensuring a comprehensive evaluation of the proposed methods.

## Conceptual Framework





*Figure 5. Conceptual Framework*

The conceptual framework illustrates the process of the application of hybridized Long Short Term Memory (LSTM) with Support Vector Machine (SVM). The framework is divided in three stages: input, process, and output. The Input stage involves the EDA signal data which is the Electrodermal Activity artifact correction Benchmark or the EDABE dataset and the gamma and c hyperparameters. The process stage involves the preprocessing of raw data, extracting relevant features, training various classifiers such as SVM, Gradient Boosted Tree, Random Forest, LSTM, and CNN, and training the hybrid LSTM-SVM. Output stage includes the preprocessed EDA signals, the selected features, performance metrics such as AUC, F1-Score, Precision, Recall, and Accuracy for both baseline classifiers and the hybrid LSTM-SVM model and the labeled artifact whether the signals is artifact or not artifacts.

## Scope and Delimitations

This study focuses on the application of hybridized Long Short Term Memory (LSTM) with Support Vector Machine (SVM) in the detection of artifacts in Electrodermal Activity signals for stress detection. The study will utilize the use of pre-existing datasets, specifically the Electrodermal Activity artifact correction Benchmark or the EDABE datasets which contain electrodermal activity (EDA) recordings of hand and body motion artifacts. This study is limited only in detection of artifacts as the removal or the correction of artifacts as with the benchmark study of Llanes-Jurado et al. (2023) will be omitted. The process of this study includes the collection as well as the preprocessing electrodermal activity (EDA) data, training the hybridized LSTM-SVM using the processed data to differentiate between authentic physiological signals and artifacts, and finally identifying artifacts from novel electrodermal data using the trained model.

## Significance of the Study

This study holds significance in the improvement in Electrodermal Activity (EDA) for stress detection. This study will be beneficial for the following:

**Caretakers.** This study will allow caretakers to potentially be able to identify and distinguish what is noise from what is a real stress response.

**Healthcare Professionals.** This study may help healthcare professionals under the pressure of time in ambulatory situations to easily interpret and distinguish what is noise/artifacts from the stress response of a patient.

**Healthcare Technology Companies.** This study may grasp this advancement to develop more precise, non-surgical stress detection devices, leading to better stress management solutions and ultimately to improve patient outcomes and quality of life.

**Future Researchers**. This study will help future researchers to build on this hybrid approach to enhance the accuracy of data that leads to more reliable result and to overall performance of EDA-based stress detection, potentially leading to more advanced and practical applications in health monitoring, psychological assessment, and stress monitoring devices

## Definition of Terms

**Ambulatory Settings** - it refers to medical services performed on an outpatient basis, without admission to hospital or other facility. Ambulatory setting is also defined as any environment where patients receive healthcare services without being admitted to hospital. This includes clinics, doctor’s offices and home-base care environments where EDA signals might be collected.

**Artifact/Artefact** - In this study it refers to unwanted disturbances or noise in EDA data that can compromise the accuracy of the measurements. It can arise due to various factors including motion, quantization errors, sudden changes in EDA associated with movement, or other sources of interference. As any anomalies in the EDA signal data that do not originate from the physiological responses but from external or technical sources. These need to be detected and filtered out to ensure the accuracy of the stress detection.

**Area Under Curve (AUC) -** it is a performance metric used to assess the effectiveness of the artifact detection model. AUC indicates better model performance in distinguishing between true artifacts and genuine physiological signals in the EDA data.

**Electrodermal Activity** - pertains to the skin’s electrical properties that shift in response to sweat secretion. These changes are frequently associated with mental and physical arousal. EDA is measured as the variation in the electrical conductance of the skin over time, reflecting the physiological arousal related to stress. The signals are collected using sensors and analyzed to detect stress levels.

**Electrodermal Activity artifact correction BEnchmark (EDABE) -** refers to a dataset and evaluation framework used to assess the performance of various artifact detection and correction algorithms in electrodermal activity (EDA) signals.

**Convolutional Neural Network (CNN) -** a deep learning algorithm designed to automatically and effectively identify artifacts in electrodermal activity (EDA) signals. CNN processes EDA signal data through multiple layers of convolutional filters to extract relevant features and patterns indicative of artifacts.

**Gated Recurrent Unit** - is a type of recurrent neural network (RNN) that was introduced by Cho et al. in 2014 as a simpler alternative to Long Short-Term Memory (LSTM) networks. GRU is used as a component of the hybridized model for processing time series EDA data, offering a simpler yet efficient way to capture dependencies in the sequential data.

**Long-Short Term Memory (LSTM)** - is a type of deep neural network that is designed to capture historical information of time series data and is suitable for predicting long-term nonlinear series. LSTM are employed to analyze and predict patterns in EDA signals by retaining information over longer periods, which helps in identifying stress related changes in the data.

**Microsiemens (μS) -** It is the unit of measurement for skin conductance, which is a key component for EDA. This study’s goal is to use a machine learning approach to filter out noise and improve the reliability of stress detection from EDA signals.

**Non-Stationary** - refers to a process of time series of statistical properties, such as mean, variance, and autocorrelation that change over time.It can also arise due to trends, cycles, abrupt changes, or other time-varying behavior in the data. This involves identifying and addressing the time-varying nature of these EDA signals when detecting artifacts.

**Psychophysiological Signals -** refers to physiological responses of the human body that are influenced by psychological factors. These signals include EDA, heart rate, respiration rate and blood pressure but it is not limited. It specifically focuses on EDA signals, which are measurements of the skin’s electrical conductance caused by sweat gland activity. EDA is closely related to the sympathetic nervous system’s activity and is often used as an indicator of emotional arousal or stress detection.

**Recurrent Neural Network** - is a deep learning model that is trained to process and convert a sequential data input into a specific sequential data output. RNNs including their variants like LSTM and GRU, are utilized to handle the sequential nature of the EDA data, enabling the detection of stress over time.

**Stationary** - It is the contrast of nonstationary processes, where the statistical properties such as mean, variance and autocorrelation, do not change over time. By identifying the segments of the signal that deviated, this can pinpoint the artifacts. Also, by analyzing its properties of the signals this can differentiate between normal variations and stress-induced changes.

**Support Vector Machine (SVM) -** it is a supervised machine learning used in classifying electrodermal activity (EDA) signals using features categorized into statistical, auto-regressive, and time-frequency related features to identify artifacts or non-artifacts.

# CHAPTER 2

# REVIEW OF LITERATURE AND STUDIES

## Stress Detection using Electrodermal Activity

Stress detection is a vital aspect of maintaining mental and physical well-being, particularly in our fast-paced modern world. Involving the identification and assessment of physiological, behavioral, and psychological indicators that signify the presence of stress in an individual. Stress detection through Machine Learning (ML) represents a groundbreaking fusion of technology and mental health, changing how we can understand and manage stress. By harnessing vast amounts of data by understanding markers such as heart rate, stimuli, breathing patterns, emotional responses, and of course electrodermal activity, the use of ML identifies these subtle patterns and signals that indicate stress level, enabling early intervention and personal health support.

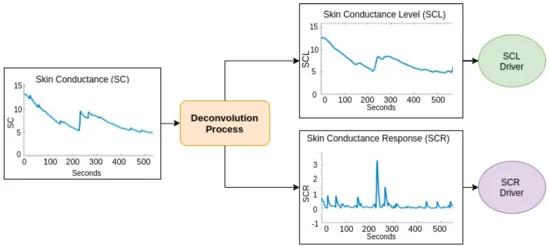
A recent systematic review by Sanchez-Reolid et al. (2022) details this as they describe how Arousal classification pipelines are set up particularly those that involve stress. Arousal, as they define, is a general physiological and psychological activation of an organism, varying on a continuum from deep sleep to intense excitation. Arousal encompasses a wide terminology, as the construct arousal is a term that corresponds to the level of cortical activation that is regulated by the ascending reticular activation system, which varies from a level of over-activation, as in the case of intense emotions or alert states, to at best attentional level for intentional action, or to levels of under-activation, as in the case of relaxation or sleep states. Because stress is closely related to arousal in many works it is strongly associated with the use of the terms distress (negative stress) and eustress (positive stress) as Le Fevre et al. (2003) states and in addition stress is considered to be a form of negative affect arousal according to Boucsein (2012). Many researchers such as Moruzzi et al. (1947) and Posner et al. (2005) agree that variation in arousal (stress being one) correlates with increases in many physiological variables such as heart rate, electrodermal activity (EDA), breath intervals and skin temperature, among others.

There are numerous physiological variables that have been used for arousal detection and its applications. EDA, according to Boucsein (2012), is considered especially useful in assessment of the arousal level due to its connection with the sympathetic nervous system (SNS). Alterations in the state of activation are unequivocally reflected as variations in skin perspiration, which affects the conductivity (conductance) of the skin. In this respect, many causal models are used to infer sympathetic activation (arousal) from EDA signals such as curve fitting, inverse filtering, general linear model for evoked skin conductance response (SCR), non-negative deconvolution, continuous deconvolution, dynamic causal model (DCM) for anticipated SCR and DCM for spontaneous fluctuations as Bach & Friston (2005) writes. As previously mentioned, there are many physiological signals or variables that can directly be collected from the human body, these being Electrocardiograph (ECG) which describe the change in heartbeat and pattern of beating, Electromyograph (EMG) which describes changes in neuromuscular activity, Electrooculography (EOG) which describe eye movements, Electroencephalography (EEG) which describes the variation of electrical signals produced in different areas of the brain, and Electrodermal Activity (EDA) which as mentioned can be used to check the arousal of an individual given that this is an important variable for measuring one's emotional state.

How EDA signals are specifically used for stress detection generally involves three stages. Firstly acquiring of course the EDA signals through a transducer or a device that measures an individual's Electrodermal Activity. The measurement of EDA signals is usually conducted in two separate ways. Through exosomatic or endosomatic means. These measurements are composed of the convolution of two signals: the electrodermal level (EDL) or the tonic component of the EDA signal which varies slowly, and the electrodermal response (EDR) or the phasic component of the EDA signal which varies rapidly. In addition, EDR, the phasic component of EDA is closely related to the activity of the sweat motor system, which is strongly associated with the sympathetic nervous system at the same time according to Bartolomé-Tomás (2020) which as Chu et al. (2024) also states is activated through stressful situations whether it be environmental or psychological and is largely in part responsible for us either avoiding it (flight) or facing it head on (fight) as a response to such a situation as Boucsein (2012) has written. To expand further the exosomatic way of measuring EDA is done by obtaining the variation of the resistance or conductance of the skin by injecting a small current onto it. This moreover is a measurement that is composed of two groups, AC and DC, depending on whether alternating or direct current is injected into the skin between the electrodes; more specifically the use of a direct current is what enables more specific EDA signals like Skin Conductance (SC), and Skin Resistance Response (SR) to be measured.

The next stage is pre-processing, which eliminates all the defects that have caused interference during the acquisition process. Here artifacts are removed and the signal is filtered, thus making it softer and eliminating noise. Pre-processing as Sanchez-Reolid et al. (2022) review puts, includes three different steps: signal normalization, detection and elimination of artifacts and filtering of noise. The researchers also highlight the importance of how artifacts that interfere with the signal must be removed, this is because a motion artifact (MAt) degrades signals very quickly and makes them unusable as Zhang et al. (2017) details. Although methods of eliminating artifacts by deflecting the signal through various softening filters can be applied, this procedure as Chen et al. (2015) and Hernandez et al. (2011) argues, causes in most cases a loss of information in EDA signals. This is the reason it may be imperative that other methods of detecting and removing artifacts such as those that are ML based be considered.

Furthermore, because EDA signals such as SC are initially structured in a way that the phasic and tonic components are entwined/convolved together, a process of deconvolution/decomposition as an important stage of signal processing, must be done in order to obtain the components from these measurements. The most common of these methods in which EDA-based literature have mostly referred to are continuous decomposition analysis, which decomposes SC data in continuous tonic and phasic activity, and discrete decomposition analysis, which separates the SC data in a tonic component and discrete phasic components with a no-negative deconvolution. In addition Convex Optimisation on EDA signals (cvxEDA) is another method of deconvolution that represents the SC as the composite of three terms: the phasic component, the tonic component and an additive white Gaussian noise that incorporates the model’s prediction errors as well as measurement errors and artifacts.

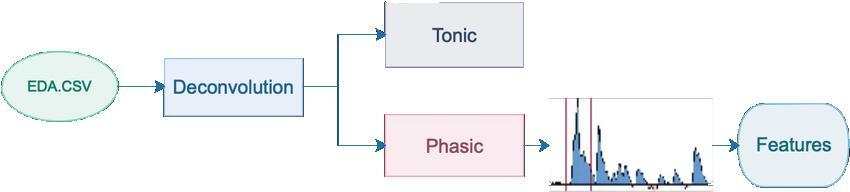


*Figure 6. EDA Signal (SC) decomposition/deconvolution from Sanchez-Reolid et al. (2022)*

Once the SC is deconvolved, the phasic component and the tonic component of the signal called the Skin Conductance Response (SCR) and the Skin Conductance Level (SCL) respectively are then obtained, and within the context of clinical analysis, the SCR which as mentioned is a physiological variable under the phasic component of the SC signal is what commonly provides the features used in EDA-based studies to provide valuable information for many scientific research fields as Posada-Quintero & Chon (2020) writes. Benedek & Kaernbach (2010), Ellaway et al. (2010), Greco et al. (2016), and again Posada-Quintero & Chon (2020) further reinforce this concept since SCR has been used to assess pain, schizophrenia, peripheral neuropathy, and more importantly stress, since the pathways of the sympathetic nervous system activity are expressed as a change in the electrical properties in SC.

Once deconvolved the SCR signal from here can be used to extract a number of features in order to perform numerous tasks of arousal classification using Machine Learning/Deep Learning models, more specifically those that involve detecting stress. The systematic review by Sanchez-Reolid et al. (2022) highlights this as the 5 most common categories of features found in EDA-based studies that used the signal as basis wherein to extract features were those relating to time, frequency, statistics, morphology (which involves the quantification of the shape of the signal), and time-frequency (which characterize the signal in time and frequency domains simultaneously).

An earlier study conducted again by Sanchez-Reolid et al. (2019) demonstrates this by their use of Support Vector Machines in measuring an individual’s Stress response using electrodermal activity. Their experiment began by recruiting a total number of sixteen participants and measuring their EDA, Heart Rate or Inter-Beat (HR), Skin Temperature (TMP), Blood Volume Pulse (BVP) using the Empatica E4 wristband which measures such variables in an individual. A group of images from the International Affective Picture System (IAPS) image library were then shown to the participants which is and was used to evoke a feeling of calmness or stress in the participant as Lang et al. (2005) instructive manual has stated. How this works is that different batches of images that have similar values of valence, arousal and dominance are selected and shown to the participants. The higher the arousal and the lower the valence (HL) was in an individual the more it indicated that he/she was in a state of calm. Conversely when the arousal was lower and the valence was higher (LH) the more it indicated that he/she was in a state of stress. Furthermore, the system then starts to randomly repeat each of the image blocks that elicit both HL and LH conditions, and when each block is finished, the Self-Assessment Manikin (SAM) questionnaire which is used to enable individuals to rate how pleasant/unpleasant (valence), calm/excited (arousal) and controlled (dominance) they felt when looking at the images, was answered. This was largely the annotation process as the signals that were recorded during image showing are synchronized with the events related to the occurrence of each batch of images. This synchronization will help to detect calmness and stress conditions as this study by Sanchez-Reolid et al. (2019) details.



*Figure 7. Flowchart of feature extraction from EDA signals from Sanchez-Reolid et al. (2019)*

In this study three categories of these kinds of features were used for their stress detection model (SVM): time-domain, morphological, and frequency based features. Under time-domain based features mean value (M), standard deviation (SD), maximum peak value (MA), minimum peak value (MI), dynamic range (DR), mean of the first order derivative (FM), standard deviation of the first order derivative (FSD), mean of the second order derivative (SM), and standard deviation of the second order derivative (SSD) of the SCR signal were all calculated in windows/segments of both 4 and 20 seconds. Under morphology-based features arc length (AL), integral area (IN), normalized mean power (AP), perimeter and area ratio (IL), energy and perimeter ratio (EL), skewness (SK) and kurtosis (KU) of the SCR signal were all calculated in windows/segments of both 4 and 20 seconds. Lastly, under frequency-based features the fast Fourier transform (FFT) through bandwidths F1(0.1,0.2), F2(0.2,0.3) and F3(0.3,0.4) were calculated over the obtained SCR signal.

Once the features were obtained windows/segments of both 20 seconds and 4 seconds were given as input to the SVM model. Conclusively this gave as Sanchez-Reolid et al. (2019) writes good results as their SVM models (which used fine-gaussian, linear, cubic kernel functions respectively) gave 5, 7, and 10-fold cross validation accuracies of 70.8%, 75%, and 70% respectively using a 20 second window. Similarly using a 4 second window their SVM models gave 5, 7, and 10-fold cross validation accuracies of 87.7% for all 3 kinds of folds; these models used quadratic, linear, fine gaussian, and cubic as their kernel functions.

In a similar study by Albertetti et al. (2021) where they also sought to use Physiological Signals like EDA for stress detection, their experiment akin to the study conducted by Sanchez-Reolid et al. (2019) also involved participants subjected to a stressful period that ranged from emotional arousal, intellectual efforts, physical exercises, and pain. And after each stressful period exposed to the participants a minimum 7 min of relaxation with some peaceful music and videos of nature followed thereafter. Once done, each participant was asked to report their perceived stress level which consisted of 4 options (1) I feel more relaxed, (2) No difference (3) I feel less relaxed, and (4) I feel more stressed. Hence, labeling and annotation of the signals were in large part achieved through the participants' relative changes of their perceived stress levels after being subjected to the numerous activities involving stressful conditions. Once the signals were acquired, decomposition of the EDA signal to obtain the phasic component or the SCR signal and the tonic component or the SCL signal were applied, as the SCR as previously established is more closely associated with activation of the SNS when faced with stimuli such as stressors. Features extracted in this study consisted of time-based and frequency-based features, which calculated the SCR amplitude peak counts, the minimum value found in the section/segment, the maximum value in the section/segment, area under curve, mean of first order derivatives, mean of negative values of first order derivatives, and Hjorth features [19] and the energy of the signal, summation of FFT harmonics, area under curve of FFT, standard deviation of FFT, mean of FFT, and signal values in the frequency domain with respect to each feature domain. As the researchers used Macro F1-Score and AUC (Area Under Curve) to gauge the performance of their proposed models these being Decision Trees, Recurrent Neural Networks (RNNs), Convolutional Recurrent Neural Networks (CRNNs), these said models over multiple windows had achieved the highest scores of 58%, 71%, and 65% respectively, showing that the use of an RNN based method achieved the highest Macro F1-Score in terms of detecting stress using EDA signals.

While the experiments conducted by Sanchez-Reolid et al. 2019 and Albertetti et al. (2021) has shown promising results in detecting stress using physiological signals such as EDA, such signals tend to still be corrupted with noise/artifacts, hence the challenges encountered by previous research and thereby the researchers, are all the more reasons to consider more robust and efficient ways of detecting artifacts such as those that involve Machine Learning. Although Albertetti et al. (2021) argues that that other physiological signals such as HR (Heart Rate), HRV (Heart Rate Variability) in contrast, are less reliable in long terms studies since they are often heavily corrupted with the movement artifacts, they also consider EDA signals to be affected by this problem. Sanchez-Reolid et al. (2019) moreover emphasizes this as in using EDA signals for stress detection, a number of limiting factors must be considered. This most importantly revolved around the quality of the data obtained. Because, in acquisition systems based on physiological signals, it is common that artifacts that damage or worsen the signal appear, hence, it is of utmost importance to avoid artifacts and events of disconnection of the electrodes with the skin. In addition Zhang et al. (2017) also emphasizes the disadvantages of these artifacts as motion artifact (MAt) degrades signals very quickly and makes them unusable. Taylor et al. (2015) moreover has detailed that continuous and unobtrusive measurement of EDA using wearable devices makes the signal collected vulnerable to several types of noise/artifacts. If these artifacts remain in the signal when it is analyzed they can easily be misinterpreted and skew the analysis; for example, they may be mistaken for a skin conductance response (SCR) (a physiological reaction that may indicate increased stress). Although methods of eliminating artifacts by deflecting the signal through various softening filters can be applied, this procedure as Chen et al. (2015) and Hernandez et al. (2011) argues, causes in most cases a loss of information in EDA signals. This is why it may be imperative that other methods of detecting and removing artifacts such as those that are ML based be considered, as the enablement to detect these problems in time will help to improve the rest of the process of stress detection. As this section established the promise of stress detection as well as the challenges it faces such as the presence of artifacts the next section will focus on the literature precisely aiming to address the detection of artifacts.

In order to produce estimations of the stress level from sensor data, the following procedures has been used by Kocielnik et. al.(2013):

* + - 1. Remove the first 15 minutes and the last 10 seconds of skin conductance measurements;
      2. Remove signal affected by losing contact with skin;
      3. Identify and remove anomalies;
      4. Smooth the signal; and
      5. Define a slicing of skin stress estimation values into five arousal categories in order to ease the interpretation by the user.

In the last step, min-max algorithm is used for overlapping 5 minutes windows to

find the most calm period and maximal EDL value, denoted as . Referencing on the 300 bins histogram, they define the value as the value before the first empty bin above . Considering that the fluctuations of SC in the very calm states are low,

## Electrodermal Activity Artifacts Detection using Machine Learning Techniques

To establish context, artifacts are specifically described as "changes in the recorded biosignal that do not stem from the signal source in question", which might be created by the recording technique or recognized physiological responses in the system that are not electrodermal signals. These may result in waste of efforts in obtaining electrodermal signals since they give rise to unreliable data because of low quality signals extracted. This work used several methods for an autonomous technique to detect artifacts in the structure of EDA signals and evaluated the signal quality in terms of thermoregulation responses (user movement, and ambient temperature). The XGboost Classifier attained a recall of 98%, a remarkable gain of 42 percentage points over the baseline classifier. It is claimed that the technique can replace or minimize the efforts of human specialists to visually analyze the retrieved data, but more extensive future study into additional elements that may appear similar to genuine signals would make a substantial contribution as Gashi et al. (2020) detailed.

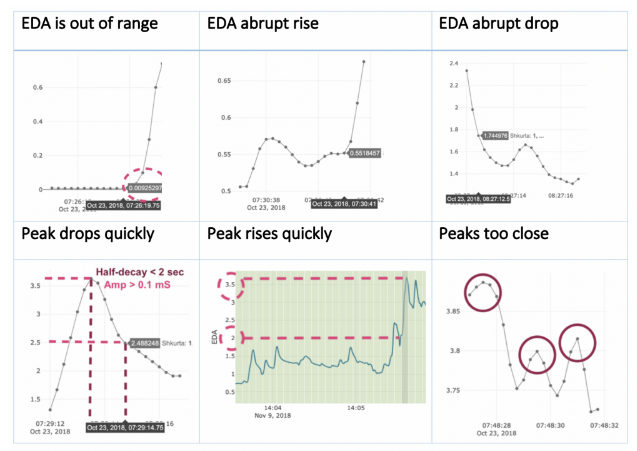
On another note how artifacts are detected by Machine Learning and/or Deep Learning models is a computationally complex process, however manual preliminary annotation and/or labeling of raw EDA signals such as in studies by Taylor et al. (2015), Hossain et al. (2022), and Llanes-Jurado et al. (2023) have all used different criteria whether or not a signal was to labeled an artifact or not an artifact. Firstly in the study by Taylor et al. (2015) a signal was labeled an artifact if firstly, the peak did not show exponential decay, (however it may be dependent on the context e.g. if two SCRs (skin conductance responses) occur close together in time, the first response may not decay before the second begins, yet this is not considered an artifact), secondly, if it has a quantization error with ≥ 5% of signal amplitude, thirdly if it has a sudden change in EDA correlated with motion, and finally that the signals SCL (skin conductance level) is less than or equal 0.

In the paper by Hossain et al. (2022) they validated their motion artifact detection algorithm, by systematically adjudicating the dataset as artifacts and clean signal albeit manually. Akin to Taylor et al. (2015) their process involved obtaining three experts/observers to manually annotate the signals using specific criteria. No fixed window was defined for the annotation, as the observers were able to mark the start and end of artifact segments. Firstly, a signal was annotated as a motion artifact if the EDA signal ranged outside 10 seconds to 40 seconds inclusively, secondly if the EDA signal changed quickly i.e. faster than ± 10, thirdly if the EDA peak decays i.e. (EDA is considered noisy if EDA peak does not follow an exponential decay except when there are two peaks within a short period of time), and finally if there was correlation of reference and noisy EDA channels (however sometimes a signal is considered clean and without artifacts if and only if the correlation coefficient was greater than or equal 0.95). In addition, in the study by Llanes-Jurado et al. (2023) although the use of experts to manually annotate the raw EDA signals according to a set of criteria was again conducted, explicit criteria such as those used in the two previous studies were not shown.

To identify and label whether the signal received from the sensor is a genuine EDA response or an artifact, here are four criteria as guidelines for annotating the signal to label if it's an artifact according to Hossain et. al. (2022):

| **Index** | **Criteria** |
| --- | --- |
| 1 | EDA out of range (*EDA range -10s to 40s*) |
| 2 | Quick change in EDA (*if EDA changes faster than* ) |
| 3 | EDA peak decays (*EDA is considered noisy if EDA peak does not follow an exponential decay except when there are two peaks within a short period of time*) |
| 4 | Correlation of reference and noisy EDA channels (*considered clean only if the correlation coefficient is* ) |

*Table 1. Guidelines for Annotation of EDA Signals*

*Figure 8. Some guidelines developed to aid human annotators in the data labeling process from the study of Gashi et. al. (2020)*

## Limitations of Machine Learning and Deep Learning Techniques in EDA Artifact Detection

Similarly, Llanes-Jurado et al. conducted a study titled "Automatic Artifact Recognition and Correction for Electrodermal Activity in Uncontrolled Environment". al. (2021), wherein the researcher discussed the influence of movement artifacts on recorded EDA signals in uncontrolled circumstances, resulting in the obscurity of significant patterns. This study investigated the application of a variety of machine learning and deep learning technologies, including support vector machines, recurrent neural networks (RNNs), and convolutional neural networks (CNN). In the experiment, the model that employed an RNN fed with the raw data recognized 72% of the artifacts and achieved an 87% accuracy rate. The identified artifacts were then automatically corrected using linear interpolation and a high degree polynomial. When assessed, the automatically and manually adjusted signals showed variations from the raw signals. The study produced significant results with regards to the future experiments that may be able to improve and develop artifact detection in EDA signals.

In an earlier study by Taylor et al. (2015), they at the time also sought develop not only a classification algorithm for automatically detecting artifacts, but an online system hosted at eda-explorer.media.mit.edu that will apply the algorithm to users’ uploaded EDA files in order to provide them with an analysis of which portions contain artifacts. Focusing however on the development of an artifact detection model their research worked on the collection of EDA signals from 32 participants and with these signals they built an interface that allowed their two experts to review these epochs and assign a label of either ‘artifact’ or ‘clean’ based on a set of criteria that were defined. Although the classification labels were created using these criteria, the website also provides the ability for other researchers to agree to label their own data according to their individual application needs. Once signals were annotated/labeled by the experts, statistical and time frequency related features were then hand crafted and engineered manually to subsequently feed into their proposed set of models for ensemble training. Wrapper feature selection was implemented to reduce redundant features from the set of hand crafted features. The data was then split into training, validation and testing to feed into a variety of algorithms such as Neural Networks, Logistic Regression, Random Forest, Naive Bayes, K-Nearest Neighbor, and Support Vector Machine, with default hyperparameters. The model that produced the highest validation accuracy was selected which ended up being the SVM, as such it was focused upon for the remainder of the paper and in which produced an overall training, validation, and testing accuracy of 76.0% 96.95% 95.67% following a hyperparameter combination that used a Radial Basis Function as its kernel, a β value of 0.1, and a C value of 1000, after prior tuning of the kernel, β, and C hyperparameters. Overall SVM proved to be a promising algorithm given the high validation and testing accuracy it had.

Hossain et. al. (2022) also explored the automatic detection of electrodermal activity data using machine learning. In this study, the researchers worked on an annotated electrodermal database to label data as clean or noisy using a reference signal without motion artifact for a more accurate resolution. In the methodology of the study, a binary classification to detect the EDA segments with motion artifacts was developed. Features from the signals were then extracted, classified, undergone hyperparameter tuning, and evaluated. This experiment resulted in automatic detection of motion artifacts using a 5-fold cross validation method with the Gradient Boosted Tree out performing all other classifiers proposed (including the baseline models originating from Taylor et al. (2015) and Kleckner et al. (2018)) in terms of Accuracy, F1-Score, Sensitivity with mean scores of 94.82%, 91.61%, 92.90% respectively, excluding however Specificity since the LDA (latent dirichlet association) model had a mean score of 97.20% but nevertheless still making the Gradient Boosted Tree a close second having a mean score of 95.75%. However, there is still room for further studies as this study had the limitations of the data being collected from only 20 subjects which may not reflect the overall population and hence, might not be sufficient to characterize most typical types of motion artifact.

In the study of Sánchez-Reolid et al. (2022), they utilized machine learning techniques for arousal classification from electrodermal activity (EDA), employing various methods to handle the complexities of EDA signal. Using first ML techniques, Support Vector Machines (SVMs) having different kernels, next is Auto-HIdden Markov Models (AHMMs) for temporal modeling and Discriminant Analysis (DA) for dimensionality reduction were key techniques. Decision Trees (DTs), including ensemble methods, and Naive Bayes methods were chosen for their simplicity and robustness. Logistic Regression (LR) was used for binary classification, while K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANNs) tackled EDA signal complexities. The artifact detection process involved signal normalization, artifact removal with filters, noise reduction, feature extraction, and model evaluation using accuracy, precision, recall, specificity, F1-score, AUC, and ROC metrics, ensuring robustness against artifacts and noise.

On the other hand, Tronstad et al. (2022) explore various time series data techniques for extracting data in Electrodermal Activity analysis. Time series data plays a pivotal role in modeling and analyzing where the interplay between cognitive processes and electrodermal responses, employing methods such as linear time-invariant (LTI) systems to discern the impact of stimuli on EDA and infer cognitive or neural inputs from the data. Additionally, time series data is integral to deconvolution schemes, aiding in the estimation of sympathetic nervous activity (SNA) time series, peak scoring for identifying physiological or psychological responses, statistical analysis to discern patterns in interpulse intervals, and model inversion for inferring cognitive or neural inputs from observed EDA responses.

A study by Llanes-Jurado et.al. (2023), utilized a dataset called Electrodermal Activity Artifact Correction Benchmark (EDABE) dataset comprising raw EDA signals and manually corrected signals as ground truth, collected from 43 participants in a VR study inducing stress. Four models were then developed and trained, two of which were approached in a manner that used traditional machine learning methods by Taylor et al. (2015) and Hossain et al. (2022). Llanes-Jurado et al. (2023) had essentially replicated the same methods used in the paper by Taylor et al. (2015) which employed the extraction of 62 hand crafted features in total (40 of which was chosen using a backward feature selection method based on a Support Vector Classifier), then fed to three different classifiers: Logistic Regression, Support Vector Machine (SVM), and a Random Forest Classifier. Subsequently, in order to select the best model for each classifier, Llanes-Jurado et al. (2023) used hyperparameter tuning for each category to select the best model out of each classifier. The model with highest validation accuracy was selected among other fitted/trained models of each of the three kinds of classifiers. On a similar note, the researchers replicated also the same methods used in the more latter paper by Hossain et al. (2022) which employed the extraction of 50 hand crafted features in total (40 of which was chosen using a different backward feature selection method mainly a Random Forest Classifier), then fed to three different classifiers: SVM, Gradient Boosted Tree, and a Logistic Regression Classifier. Again, in order to select the best model for each classifier, the researchers used hyperparameter tuning for each classifier category to select the best model. The model with highest validation accuracy was selected among other fitted/trained models of each of the three kinds of classifiers. Replicating the Taylor et al. (2015) and Hosasin et al. (2022) methodologies the models that emerged having the highest performance in terms of accuracy was the Random Forest (contrary to the SVM model by Taylor et al. (2015)) and the Gradient Boosted Tree classifier respectively. These two baseline models were now compared to their proposed LSTM-CNN as this was projected to be able to identify temporal evolutions and features even within raw EDA signals itself. It had the following hyperparameter configuration: first two layers were LSTM layers of 16 neurons that returned the hidden state output for each input time step. Subsequently, the network included four convolutional levels, each of which featured three convolutional layers with a batch-normalization operation performed after each convolution. Finally, each level included a dropout value of 0.05 and a max-pooling operation of size 2. The numbers of filters in each level were 32, 64, 128, and 256; kernel size was 5. Finally, the model featured two fully connected layers of 256 and 16 neurons and a final fully connected layer comprising a single perceptron with a sigmoid activation function. The model was trained with the rmsprop optimizer at a learning rate of 5 × 10−5 and a batch size of 16. Due to the imbalance of the dataset labels, the cost function used to train the model was the Dice-Sørensen coefficient (DSC). The model had an early stopping threshold of 30 epochs. The percentage of artifacts in the training set was 12.60%. No filter was applied to the raw signal and for each 5 s segment, min–max scaling was applied. After training, testing the LSTM-CNN proposed by Llanes-Jurado et al. (2023) the model had its AUC (area under curve), Dice Sorensen Coefficient (F1-Score), TPR (True Positive Ratio), and Kappa mean scores of 76%, 57%, 65%, and 49% respectively essentially out performing all other baseline models by Taylor et al. (2015) and Hossain et al. (2022) highlighting the ability of the LSTM to extract temporal features from even raw EDA signals without having to hand craft features from the dataset itself. Although the proposed model was third and a close second in terms of the Accuracy and True Negative Ratio metrics, this was negligible, as due to the present imbalance of the dataset, AUC, DSC, TPR, Kappa, are better overall metrics as opposed to solely just relying on an Accuracy metric itself, because of an imbalance of labels in the dataset. Overall despite skewed labels the proposed model represented a large increase in artifact detection performance relative to previous methodologies, being able to detect artifacts 59.88% of the time given a 50% threshold for the loss function, and 81.39% of the time given a 20% threshold for the loss function.

Although previous studies particularly those using traditional machine learning methods did have promising results such methods still do have limitations in healthcare signal processing. A study by Sun et al. (2018) found that hand-designed EEG feature extraction methods resulted in poor analytical performance. They addressed this by using recurrent autoencoders for feature extraction.

## Advantage of LSTM

Hussein et al. (2018) further supported this by employing LSTM networks to capture high-level representations in EEG signals. In their approach, a fully connected layer was used to extract robust, epileptic-relevant features, while a softmax layer provided predicted labels. This method maintained high detection performance, particularly in identifying artifacts like eye movements, muscle movements, and background noise. The results on a benchmark clinical dataset reveal the prevalence of the proposed approach over the baseline techniques; achieving 100% classification accuracy, 100% sensitivity, and 100% specificity. Additionally while EEG signal data can be noisy in real-life conditions the researchers proposed model proved robust, since it maintained high detection performance in the existence of common EEG artifacts (muscle activities and eye movement) as well as background noise.

Additionally, in a study by Jamshidzadeh et al. (2024), one of these limitations with regards to the use of traditional ML methods was specifically the inability of SVM models to extract significant features from data. They also added that there is always a dependency of the accuracy of the model on the choice of input combinations, which requires the use of robust input selection methods i.e. feature selection which can be prone to bias due to it being manually done by humans and so feature selection methods like SVC and Random Forest have been used as with studies by Taylor et al. (2015) and Hossain et al. (2022) in order to select appropriate features from the data.

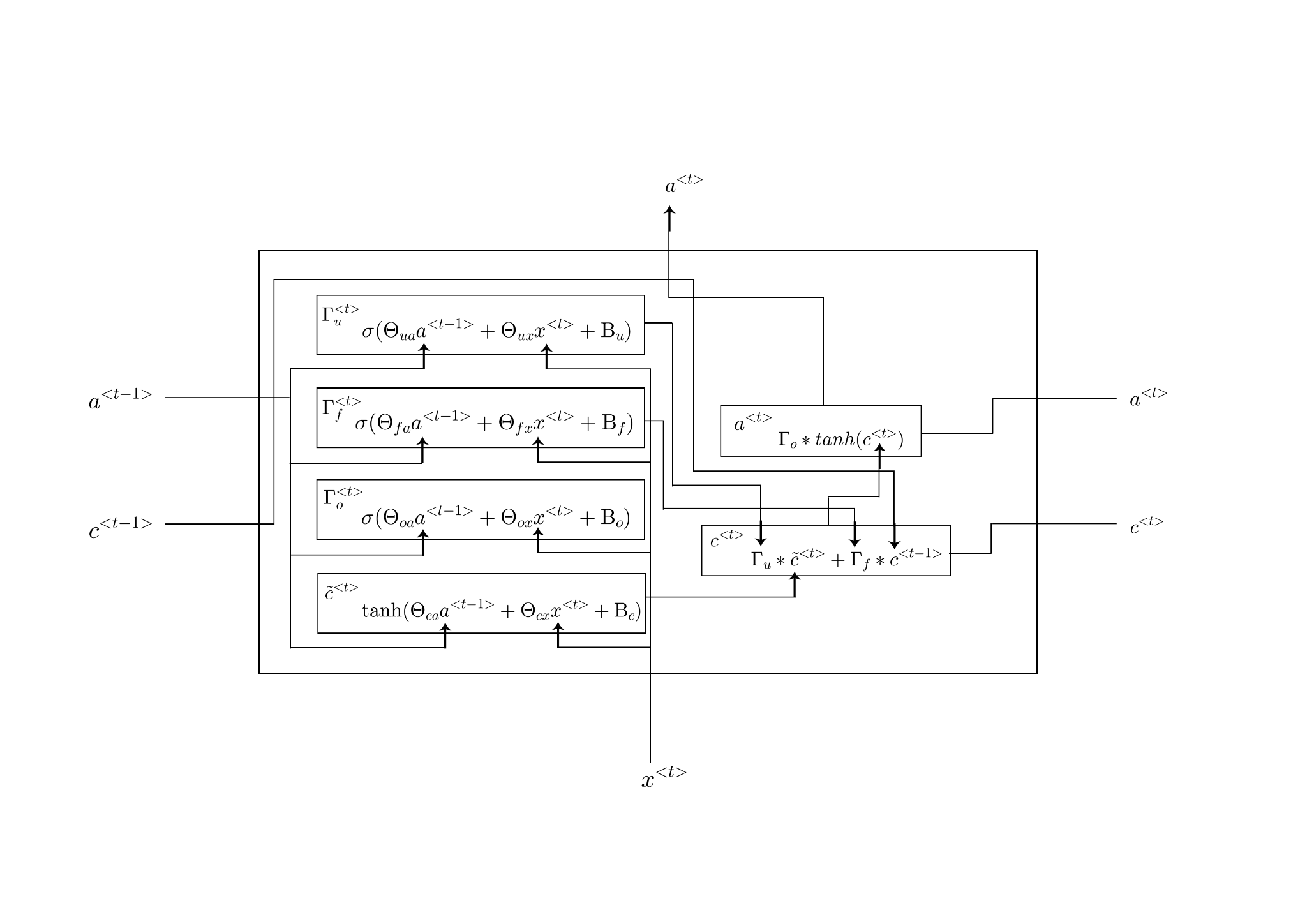
This is further reinforced in the studies by Koklu et al. (2022) and Agarwal et al. (2023) regarding water quality prediction as they state that predicting water quality, it was indeed essential to develop SVM models with thorough feature engineering and extraction methods for accurate water quality prediction due to their limitations in extracting important features from time series data, since they may not accurately handle complex data and may not perform optimally.

Moreover, with the aforementioned study of Llanes-Jurado et al. (2023), while it has been established that convolutional neural networks achieve superior performance on high dimensional data such as images and used in different automated tasks, such as classification, detection, segmentation, data augmentation Szegedy et al. (2016). CNNs, however, according to Wang et al. (2020), Park & Yang (2019), Li et al. (2022), are still considered as a ‘Black box’ in terms of the underlying mechanism which makes it difficult to interpret the results and have confidence that they provide the optimal solution. A paper by Azam et al. (2023) also states that the ‘black box’ nature of CNNs is a key research interest currently where different research has attempted to provide different insight. There are several research questions related to the ‘black box’, such as the inner mechanism of CNN layers, feature interpretation and decision making schemes and as such CNNs are still poorly understood, not only by non-technical users but also by experts. This lack of knowledge according to Lange et al. (2018) may cause ambiguity and a hesitance in relying on the predictions of CNNs, especially in critical applications like the medical domain. And by opening up the ‘black box’ it can increase the confidence of users such as medical specialists in the results of neural networks as per the studies conducted by [Ferdinand & Mercier, (n.d.](https://www.sciencedirect.com/science/article/pii/S2667305323000583?fbclid=IwZXh0bgNhZW0CMTAAAR237Nav-d71WwmVzTLNLOBqE1XXAkd6BSOqSv_C7nxX_ucXZ_Q7_5pZ-3o_aem_AQigPAAWF8bKS27hMiDqyvmg_qlmIqHGDdt3kI_JnosTqgs_apWmIkyg3SSq1gQ4LIR6basuDJs93-vtpaEE7NA7#bib0012)), [Brahimi et al. (2018](https://www.sciencedirect.com/science/article/pii/S2667305323000583?fbclid=IwZXh0bgNhZW0CMTAAAR237Nav-d71WwmVzTLNLOBqE1XXAkd6BSOqSv_C7nxX_ucXZ_Q7_5pZ-3o_aem_AQigPAAWF8bKS27hMiDqyvmg_qlmIqHGDdt3kI_JnosTqgs_apWmIkyg3SSq1gQ4LIR6basuDJs93-vtpaEE7NA7#bib0004)), and [Dependent et al. (2021](https://www.sciencedirect.com/science/article/pii/S2667305323000583?fbclid=IwZXh0bgNhZW0CMTAAAR237Nav-d71WwmVzTLNLOBqE1XXAkd6BSOqSv_C7nxX_ucXZ_Q7_5pZ-3o_aem_AQigPAAWF8bKS27hMiDqyvmg_qlmIqHGDdt3kI_JnosTqgs_apWmIkyg3SSq1gQ4LIR6basuDJs93-vtpaEE7NA7#bib0008)).

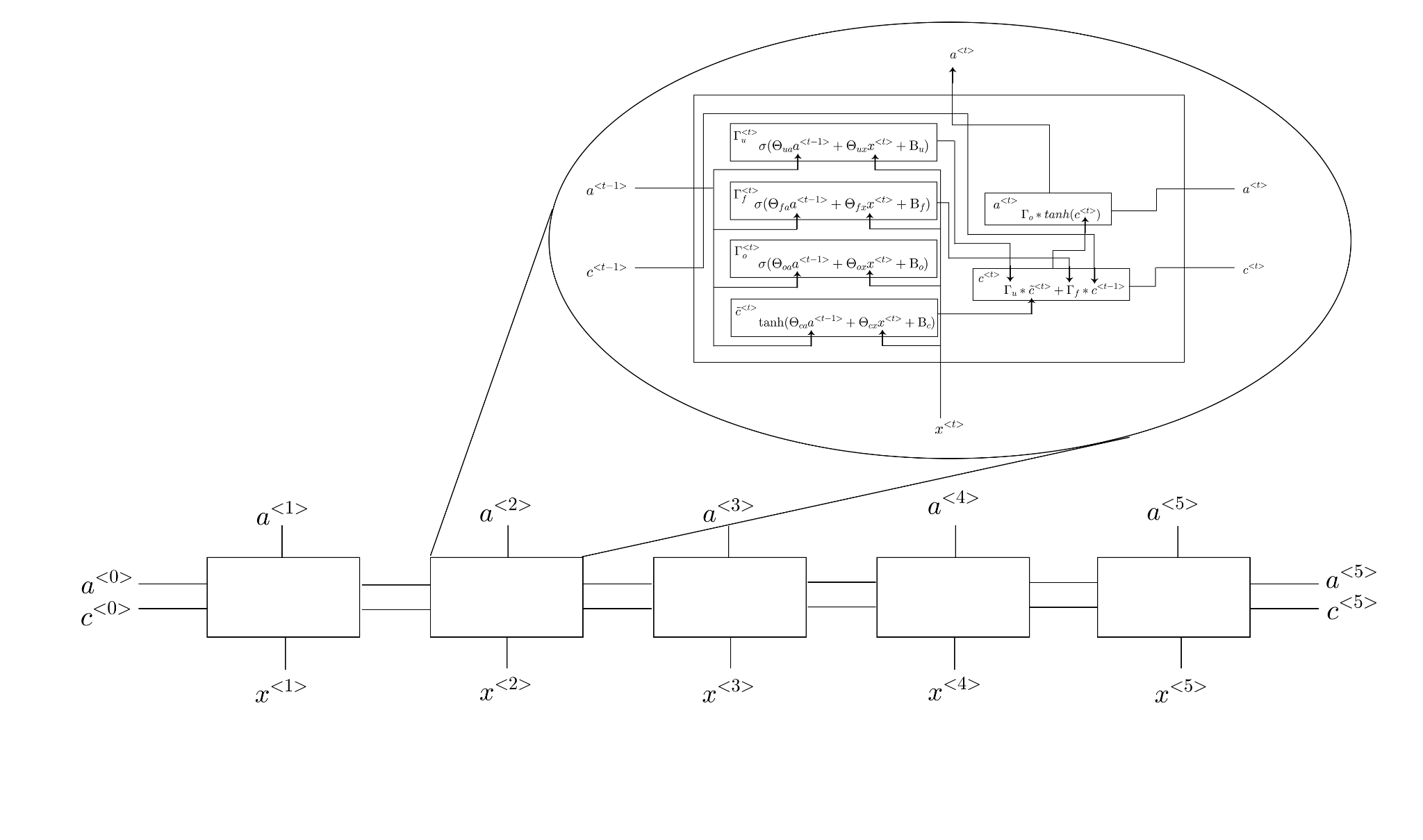
In relation to CNNs, as it is a subset of deep learning, deep learning is a more complex subset of machine learning. Deep learning models, albeit require more training time, often have provided higher accuracy as detailed by a paper by Kamath et al. (2018) because they can perform automated feature extraction and classification concurrently, whereas a feature selection process is required prior to training a machine learning algorithm.

According to Jamshidzadeh et al. (2024), Recurrent Neural Networks (RNNs) represent a specialized class of Artificial Neural Network (ANN) models meticulously designed for the analysis of sequential data. RNNs also incorporate cyclical connections within their architecture, which equip them with the unique capacity to manage and process sequences of variable lengths, as detailed by Wu et al. (2020). Fundamental to the RNN’s operation is an internal hidden state, serving as a form of memory. This hidden state undergoes continual updates at each time step and is under the influence of both the incoming input data and the hidden state from the previous time step. Conceptually, the architecture of an RNN can be envisioned as an interlinked chain of repeating modules or cells. These recurrent connections within the RNN structure facilitate the preservation of information from previous time steps. At any given time the hidden state is computed as a function of the current input at time and the preceding hidden state or at time employing a set of learned weights and an activation function. This sequential data processing mechanism culminates in the RNN’s ability to systematically process input data as it unfolds in a temporal sequence as detailed also in a paper by Weerakody et al. (2021). Specifically, the input is fused with the hidden state from the previous time step to compute the new hidden state , a process underpinned by a set of adaptively learned weights and an activation function. Following the update of , the RNN is poised to generate an output corresponding to the current time step . This description according to Jamshidzadeh et al. (2024) encapsulates the essence of how RNNs operate in the context of sequential data analysis.

However, as Hochreiter (1991) and Mozer (1992) stated, RNNs are limited to look back in time for approximately ten timesteps rendering the RNN incapable of remembering information with longer term dependencies. This is due to the fed back signal that either vanishes or explodes. Nonetheless, as detailed by Gers et al. (2000) Gers et al. (2002), Hochreiter & Schmidhuber (1997), and Perez-Ortiz et al. (2003), this issue was addressed with the advent of Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN). Unlike its latter and computationally simpler counterpart, the RNN, LSTM networks are to a certain extent biologically plausible and capable of learning more than 1,000 timesteps, depending on the complexity of the built network as O’Reilly & Frank (2006) and Hochreiter (1997) has written.



*Figure 9. Internal mechanism of LSTM cell*

**

*Figure 10. Full LSTM model*

In place of a traditional RNN cell, the LSTM according to Wan et al. (2022) and Jaseena & Kovoor (2021) replaces the self-connected hidden units of a traditional RNN with memory blocks. The memory blocks contain multiple memory cells. Each memory cell stores information over a specific period of time and can be accessed and modified by different gate units. Specifically, LSTMs have three types of gates: input , forget , and output gates , which are responsible for controlling the flow of information. The input gate regulates the flow of new information, while the forget gate determines what information should be discarded from each memory cell. Finally, the output gate determines what information should be sent to the output of the LSTM as detailed by Wan et al. (2022) and Zhang et al. (2022). The outputs of different gates including the next hidden state and cell state at timestep are calculated as follows:

## Hybridizing LSTM and SVM

In addition, further studies such as by Mohsen et al. (2023) which involved the classification of EEG signals using an LSTM and a traditional ML method the SVM classifier, shows that the LSTM provided the best performance, with a testing accuracy of 99.00%. Moreover, the weighted average precision, recall, and F1-score for the LSTM are 99.00%. The results of the SVM classifier in terms of accuracy, sensitivity, and specificity reached 91%, 93.52%, and 91.3%. Given such results they have shown that the LSTM classifier provides better performance than SVM in the classification of EEG signals. Reinforcing the idea that when it comes to feature extraction the LSTM itself gives superior results as opposed to traditional machine learning methods that require manual feature extraction and engineering.

Going back to Jamshidzadeh et al. (2024), their study has shown that when it comes to predicting water quality parameters their proposed models which consists of a BILSTM-SVM, LSTM-SVM, BILSTM, and both a standalone LSTM, and SVM had accuracy indices of 97%, 92%, 89%, 85%, and 82%, respectively, at the training level. And as such it was concluded that the BILSTM-SVM model performed better than the LSTM-SVM model because it considered both past and future observations, secondly that the hybrid model (BILSTM-SVM and LSTM-SVM) performed better than the standalone LSTM, BILSTM, and SVM models, and thirdly that the SVM model had lower Accuracy Index (AI) and R2 values compared to other models, and all other models outperformed the SVM model. Similarly at the testing level, the AI values of the BILSTM-SVM, LSTM-SVM, BILSTM, LSTM, and SVM models were 0.94, 0.86, 0.83, 0.81, and 0.78, respectively, showing again evidence that SVM models paled in comparison to LSTM models in terms of extracting features from time series data such as water quality parameters.

In another study by Juneau et al. (2021) which sought to detect foot strikes in lower extremity amputee populations since a viable model according to the researchers would provide the basis for automated stride parameter calculation and stride segmentation, which is essential for using new fall risk and health status AI models within clinical environments, their proposed models consisted of a traditional ML approach namely a Decision Tree and a Deep Learning approach the LSTM (Long-Short Term Memory) Neural Network. Their study has shown that the best performing classifiers out of the two was the LSTM (having a configuration of batch size of 64, dropout of 0.4, one LSTM layer with 100 hidden LSTM nodes, one dense layer with 50 hidden dense nodes, and a class weighting of 1:2 (label 0: label 1) which had an accuracy, sensitivity, specificity, and precision of 99.0%, 86.4%, 99.4%, and 83.7% respectively over the decision tree classifier (having the configurations maximum tree depth of 10 and class weighting of 1:20 (label 0: label 1)) which had the aforementioned metrics of 98.7%, 82.8%, 99.2%, and 78.6% respectively showing still that a Deep Learning approach may be used to achieve higher performance metrics than traditional ML methods without the task of manually engineering & extracting features as well as selecting the most important features.

On the other hand, while deep learning methods may potentially address the problem of manual feature extraction and engineering by being able to extract high level features particularly that of biomedical signal data like EDA signals, as mentioned the problem of deep learning models particularly CNNs still have the challenge of interpretability when it comes to the predictions it produces, and due to its ambiguous nature medical professionals in particular have a hard time grasping how it arrived at such a prediction. Recent studies such as those by Mordensky et al. (2023) further support this as they proposed to use predictive modeling using traditional Machine Learning methods such as Logistic Regression, XGBoost, and SVM and more complex methods such as the use of Artificial Neural Networks in the domain of estimating moderate and high temperature geothermal resource favorability in the United States. Albeit having an imbalance of classes in their dataset i.e. having more positive labels, thereby having to use alternative performance metrics particularly the F1-score (as the performance metric accommodates for class imbalance better than other performance metrics e.g., accuracy Guo et al. (2008)), their study showed that while F1 scores of the proposed models appeared poorly, normal score transform was applied to the predictions of each model, and as a result it was found that the highest performing models was the single XGBoost and the ensemble trained XGBoost model having mean normal score transformed values of 1.88 and 1.61 respectively, of all the positive labels it predicted, over other models such as single Logistic Regression (LR), ensemble Logistic Regression (enLR), single SVM, enSVM, ANN, having similar mean normal score transformed values of 1.32, 1.30, 1.24, 1.21, and 1.30 respectively. These scores have indicated the single trained XGBoost model produces the greatest distinction between known positive and unlabeled samples and that the ensemble trained XGBoost predicts the second greatest distinction between known positive and unlabeled samples. Similarly, the substantial overlap of feature importance in the F1 and ROCAUC sensitivity analyses of the single ANN suggest that this deep learning algorithm was also detrimentally affected by its complexity when faced by the simplicity of the feature data. Hence, the highly complex machine learning models may not be as appropriate for data like that from the 2008 geothermal resource assessment the researchers used as the less complex algorithms such as XGBoost.

In another study by Sánchez-Reolid et al. (2019) they demonstrate yet again the simplicity of ML methods in the domain of stress detection from EDA signal data, as they’ve shown that the simplicity of the classiﬁcation model they proposed, namely the SVM allows the system to work in the long term, as the SVM across 3 different sets of cross validation given 20 second segments of EDA signal features managed to produce at most 70.8%, 75%, and 75% accuracy, for 5, 7, and 10 folds of cross validation. Similarly 4 second segments of EDA signal features given to the SVM managed to produce at most 87.7% accuracy in the 5, 7, and 10 folds of cross validation.

In the more recent literature in the domain of sequential data processing such as NLP and signal processing, akin to the recent paper by Llanes-Jurado et al. (2023) where they proposed a hybridized LSTM-CNN model for automatic artifact detection in EDA signal data, it has fairly become common practice to adopt approaches of hybridizing Machine Learning and Deep Learning methods to making more accurate predictions from sequential data.

A study by Cimino & Dell’Orletta (2016) demonstrates this by hybridizing an LSTM and SVM in the domain of sentiment analysis i.e. subject classification, polarity detection, and irony detection. With this proposed model they sought to compare it with other baseline models, these being a standalone LSTM and an SVM model. They consider an SVM since it is an extremely efficient learning algorithm and hardly to outperform, unfortunately these type of algorithms capture “sparse” and “discrete” features in document classification tasks, making really hard the detection of relations in sentences, which is often the key factor in detecting the overall sentiment polarity in documents as Tang et al. (2015) details. The mechanism of an SVM is defined as follows as per the book by Vapnik (1995):

*Eq. 1 Linear SVM*

Where is the input feature vector at row , the non-bias weight/coefficient vector, and the bias weight/coefficient.

*Eq. 2 Cost/Loss function*

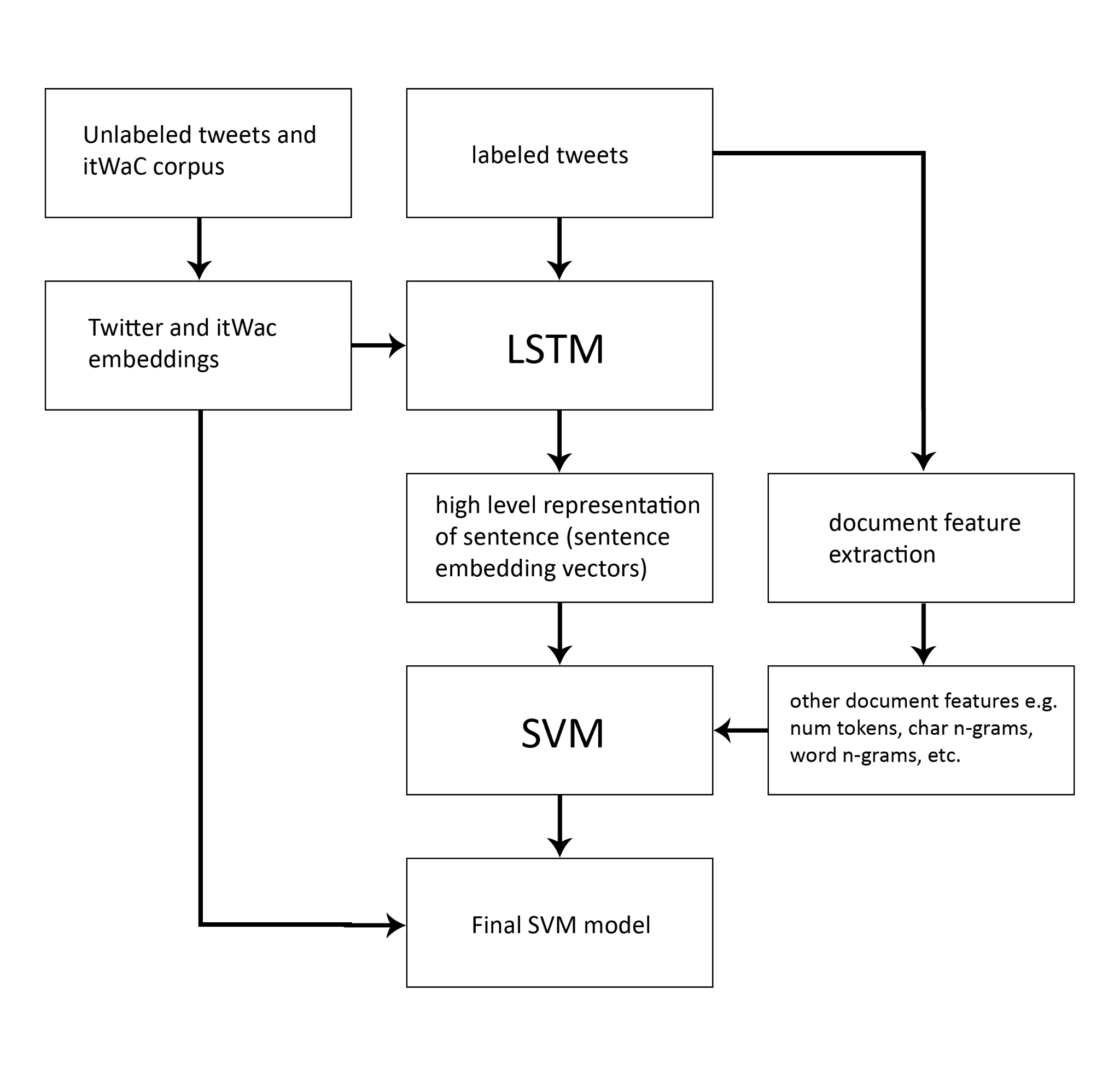
where is the penalty coefficient, m is the number of training data points, λ− i and λ+ i are the penalty for the prediction errors; and out is the observed output. Equation (1) is rewritten based on the kernel functions:

*Eq. 3 rewritten hypothesis using kernel function*

where is the kernel function. In the modeling process, the radial basis function (RBF) is widely used as a kernel function. Previous studies such as those by Vapnik (1995) indicated that the use of a RBF as the kernel function gave high accuracy.

There are many kernel functions i.e. Linear, Polynomial, Sigmoid etc. but the Radial Basis Function involves a the kernel parameter .

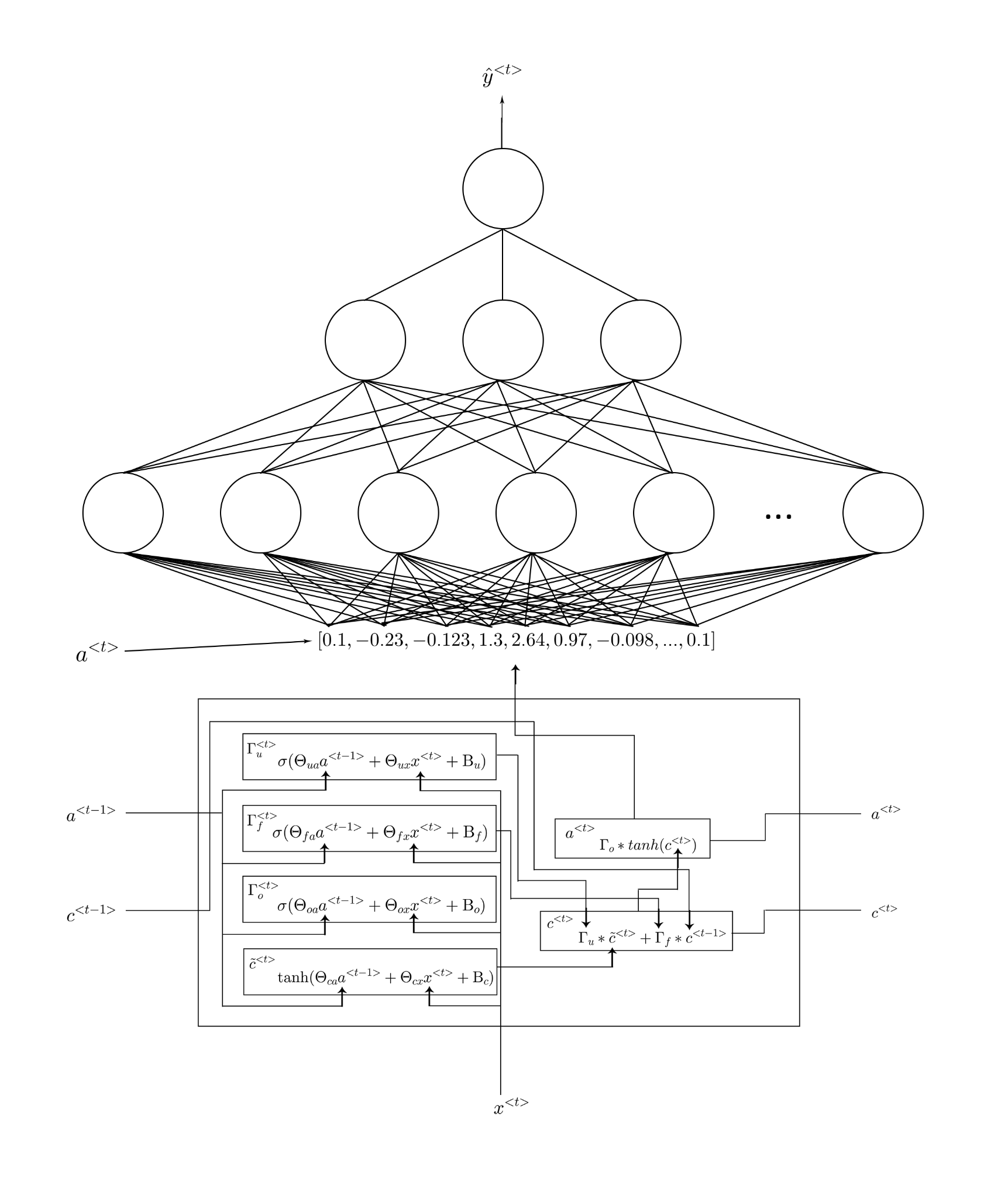
Furthermore, with the use of LSTM it was able to fortunately capture long term dependencies of a sentence, and hence in the training phase, the LSTM network was trained considering the training sentences and its corresponding gold labels of tweets in which the researchers decided to use for the task of sentiment analysis. Once the model was trained on each sentence of the training set, a sentence embedding vector consisting of real numerical values can then be obtained from the layer before the softmax classifier, which as previously mentioned would be the hidden state of an LSTM network. These sentence embedding vectors were then used as feature inputs to the SVM classifier in conjunction with a set of widely used document classification features i.e. lexical text, morpho-syntactic, and lexicon features.



*Figure 11. Cimino & Dell’Orletta (2016) hybridized LSTM-SVM architecture*

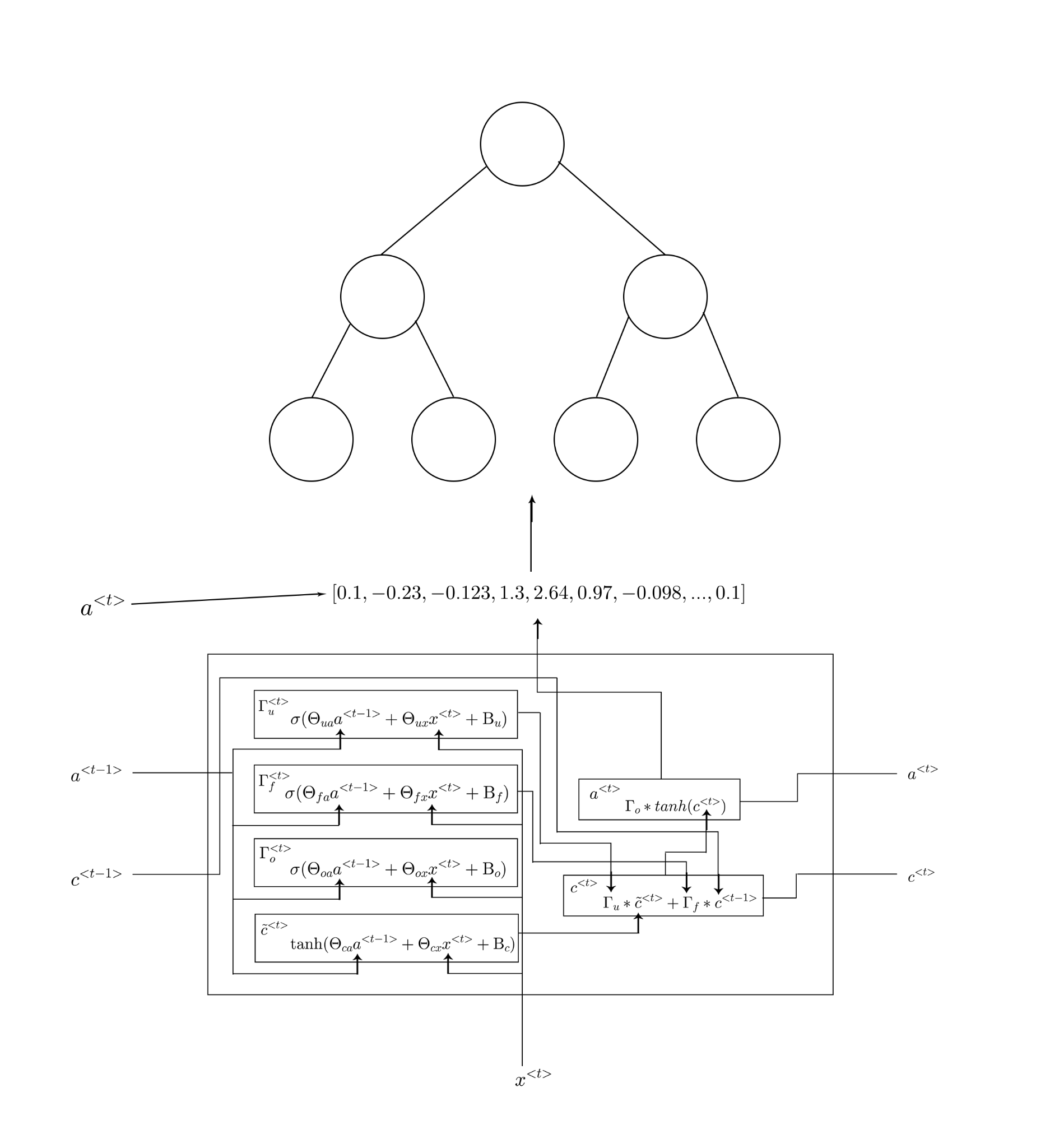
As for their results overall accuracies achieved by the classifiers on the validation data set showed that the standalone LSTM out performs specifically linear and quadratic SVM models in particular, having a 77.7%, 74.7%, and 64.6% accuracy in subject classification, polarity and iron detection tasks respectively over the latter of 74%, 73%, and 59.5% for quadratic and 72.5%, 71.3%, and 63.6% for linear. Moreover, the hybridized linear and quadratic LSTM-SVMs clearly outperformed the standalone LSTM and the linear and quadratic SVMs, achieving 76.4%, 74.3%, and 66.2% validation accuracy for linear and 78.3%, 75.4%, and 67.5% validation accuracy for quadratic for subject classification, polarity and irony detection tasks respectively.

In a more recent study Zhang & Zhang (2020) they also proposed a hybridized model specifically for urban road short-term traffic flow prediction in order to solve and analyze the problems of periodicity, stationary, and abnormality of time series to improve traffic flow prediction effect, and achieve efficient traffic guidance and control. As this was a regression task the researchers chose mean squared error, its root (RMSE), mean absolute error, and mean absolute percentage error as the performance metrics to use, in place of typical accuracy, precision, recall, F1-score etc. for classification tasks, and where the lower the error value the better the model has performed. In place of the SVM as the traditional ML method as with the previous study the researchers proposed the use of a LSTM hybridized with the XGBoost (Extreme Gradient Boosting) classifier instead, then compared with other baseline models such as CNN, a standalone LSTM, a standalone XGBoost, LSTM-RNN, a model proposed by Wang (2019), and a model proposed by Chen & Chen (2019) in order to validate their hybridized model. This was because multi-step predictions can be challenging for time series data and furthermore because fully connected layers could potentially harbor too many parameters thus leading to overfitting the traditional ML method, the XGBoost classifier was used instead as an alternative mechanism to which the final hidden state of the LSTM can be sent to, as feature vector inputs.



*Figure 12. Typical LSTM network*

Here in a typical LSTM network cell the computed hidden state is sent to a fully connected layer (dense layer) or immediately to a softmax to produce a probability vector as the prediction. However as detailed by Zhang & Zhang (2020) this dense layer will instead be replaced with the XGBoost classifier itself, where instead of the hidden state being sent to a dense layer it is sent as a representation of the samples of the training data i.e. feature vector to the XGBoost classifier.

**

*Figure 13. LSTM-XGBoost architecture*

When it came to workday forecasting, performance indices in terms of MSE, RMSE, MAE, MAPE, showed the hybridized LSTM-XGBoost model to reign superlative in all metrics having values of 7.55, 2.75, 1.86, and 4.19% respectively as opposed to other models even the second best model by Chen & Chen (2019) having values of only 9.99, 3.16, 2.03, and 10.61% respectively. Moreover, as opposed to standalone models like the LSTM and XGBoost classifier, the hybridized models prediction accuracy, namely its MAPE as opposed to the two latter standalone models, improved by 6.73% and 0.91%. Similar values were also found when it came to weekday forecasting, when it came to the hybridized model as it achieved values of 4.10, 2.03, 1.49, and 3.3% respectively for MSE, RMSE, MAE, and MAPE, as opposed to even the second best model again by Chen & CHen (2019) having values of only 4.47, 2.11, 1.58, and 7.08% respectively for the aforementioned metrics. Moreover, as compared to the standalone models of LSTM and the XGBoost classifier, the hybridized approach improved by 3.72% and 0.02%. From these results the researchers concluded that their proposed LSTM-XGBoost hybrid model not only can improve prediction accuracy, but can also perform multi-step prediction, which is an effective method for traffic speed prediction.

Finally, in a study by Agarap (2017) the use of a hybrid GRU-SVM (a much computationally simpler version of an LSTM cell) akin to that of Cimino & Dell’Orletta (2016) tandem LSTM-SVM was also implemented in order to detect intrusions and malicious activity of users in cyberspace. The research attempted to solve this problem by again replacing the Softmax mechanism of a conventional RNN and replacing it with an SVM, because conventionally like most neural networks, the RNN and its variants (LSTM and GRU) employ the Softmax activation function as its final output layer for its prediction, and the cross-entropy function for computing its loss. Furthermore, the paper also amended this norm by introducing a different loss function in place of the cross entropy loss which is typically used with the softmax activation function especially when predicting multi class labels: a margin-based loss function is used, it is defined as follows:

However because optimization of a neural network requires backpropagation (differentiation of the loss function with respect to the parameters of the network) this loss as detailed by Agarap (2017) cannot be differentiated, instead a different version of this margin-based loss function is used: the squared hinge loss, which will be differentiable and much more stable as Tang (2013) has stated.

After training and testing, results show that, although the GRU-SVM was not able to surpass the prediction rate of GRU-softmax for true negative labels having only an approximate 75.60% as opposed to the latter of approximately 95.91%, and for false positive labels having a greater prediction rate of approximately 10.70% as opposed to the latter of approximately 4.09%, the GRU-SVM outperforms GRU-softmax as it was able to predict true positive labels approximately 89.3005% of the time and for false negative labels having a lesser prediction rate of approximately 24.40% as opposed to the latter of approximately 44.39%. Moreover in terms of overall accuracy the GRU-SVM outperformed GRU-softmax in both training and testing splits with values of 81.54% and 84.15% with respect to each data split as opposed to the latter of 63.07% and 70.75% with respect to each data split. As Agarap (2017) concludes, not only did the GRU-SVM model outperform the GRU-Softmax in terms of prediction accuracy, but it also outperformed the conventional model in terms of training time and testing time. Thus, supporting the theoretical implication as per the respective algorithm complexities of each classifier.

## Synthesis of the Study

While it has been established that the use of traditional ML methods have shown promising results particularly in the domain of the detection of artifacts in Electrodermal Activity Data, as mentioned it does still have its disadvantages when it comes to the tasks of manual feature extraction and engineering and its inability to extract high level features from signal data as proven by the studies of Mohsen et al. (2023), Jamshidzadeh et al. (2024), and Juneau et al. (2021), showing that LSTMs were more suited to perform this task of feature extraction as opposed to the aforementioned methods. On a similar note, although using deep learning models for the classification of labels from signal data such as a hybridized LSTM-CNN for EDA signals proposed by Llanes-Jurado et al. (2023) also have shown promising results in the domain of the detection of artifacts in Electrodermal Activity Data, as mentioned it does still have its disadvantage of interpretability

Studies such as those by Mordensky et al. (2023) and Sánchez-Reolid et al. (2019) have suggested that when faced with the simplicity of features deep learning algorithms performance may even more so be detrimentally affected than gain benefit. Hence, much simpler data may be better solved by traditional ML methods i.e. XGBoost, Logistic Regression, SVM, etc. than highly complex machine learning models such as the aforementioned CNN. In addition CNNs as mentioned has the concern of the lack of knowledge in the processes it has done to arrive at a prediction surrounding it, and as Lange et al. (2018) argues may cause ambiguity and a hesitance in relying on the predictions of CNNs, especially in critical applications like the medical domain.

With these problems at hand the researchers propose a hybridized LSTM-SVM model. The LSTM being the mechanism that will not predict an outcome/probability but the mechanism that automatically extracts high level features from EDA signals and SVM being the mechanism that addresses the problem of complexity and interpretability of CNNs predictions, as a traditional ML method such as this may in fact address such a problem.

# CHAPTER 3

# METHODOLOGY

## Research Design

In this study, the researchers will adopt an experimental and quantitative method to achieve the objective of automatically detecting artifacts from electrodermal activity (EDA) signals used for stress detection. The study will employ hybridized Long Short-Term Memory (LSTM) and Support Vector Machine (SVM) models and compare their performance to existing models of previous researchers, Taylor et. al. (2015), Hossain et. al. (2022), and Llanes-Jurado et. al. (2023), specifically. The goal is to determine the efficacy of the hybrid LSTM-SVM model in improving the overall performance in terms of Accuracy, Precision, Recall, AUC, and F1-score, and the reliability of stress detection from EDA signals by effectively handling artifacts. To attain this objective, the researchers propose the application of hybridized LSTM-SVM to address the concern with the non-stationary characteristic of EDA data: it can change through the course of time. The study consists of independent variables which are the EDABE dataset, the existing models by Taylor et. al. (2015), Hossain et. al. (2022), and Llanes-Jurado et. al. (2023), and the proposed model, which is the hybridization of LSTM and SVM. Dependent variables comprise the performance of the baseline models and the proposed model in terms of accuracy, precision, recall, and F1-score. The researchers expect that results from the proposed model will either match or excel over the performance of the existing models.

## Sources Of Data

This study will utilize the Electrodermal Activity artifact correction BEnchmark (EDABE) dataset as its primary source of data. The EDABE dataset consists of electrodermal activity (EDA) recordings of hand and body motion artifacts. EDABE includes 74.46 hours of EDA recordings, which are influenced by hand and body movement artifacts, from 43 subjects. The dataset is split into a training set with 33 subjects (56.27 hours) and a test set with 10 subjects (18.19 hours). The recordings were obtained using a Shimmer3 GSR+ Unit at a sampling rate of 128 Hz. Each file's name contains the user ID and the expert who corrected the signal. Additionally, the file includes the signal with the following variables:

| **time** | Timestamp of the signal |
| --- | --- |
| **rawdata** | Raw data obtained by Shimmer3 GSR+ Unit |
| **cleandata** | Reconstructed clean signal performed by a human expert |
| **binarytarget** | Label of each sample as artifact or no artifact |
| **signal\_automatic** | Automatic cleaning of the signal performed by the automatic pipeline |
| **predArtifacts** | Label predicted by the automatic cleaning pipeline |
| **postProcessedPredArtifacts** | Label predicted by the automatic cleaning pipeline after post processing |

*Table 2. Signal components in EDABE dataset*

## Research Instrument

This study will be implemented using Python as the programming language. Python is a highly versatile and widely utilized programming language in data science and machine learning research, praised for its readability, simplicity, and extensive library support. Its comprehensive ecosystem of libraries and frameworks makes it ideal for implementing complex models, performing data analysis, and visualizing results.

It will also utilize the dataset, EDABE (Electrodermal Activity artifact correction BEnchmark), which comprises 74.46 hours of EDA recordings impacted by hand and body motion artifacts from 43 subjects. It is split into a training set with 33 subjects (56.27 hours) and a test set with 10 subjects (18.19 hours). The data was recorded using a Shimmer3 GSR+ Unit at a sampling rate of 128 Hz.

For the tools and technologies, the study will be utilizing TensorFlow, scikit-learn, and Google Colab. TensorFlow, an open-source machine learning framework developed by Google, is extensively used for building and deploying machine learning models, particularly deep learning models. In this study, TensorFlow can be utilized to implement and train a Long Short-Term Memory (LSTM) network, which is well-suited for analyzing time series data such as EDA signals. On the other hand, scikit-learn is a robust machine learning library in Python that offers simple and efficient tools for data mining, analysis, and model evaluation. In this study, scikit-learn will be employed to implement the Support Vector Machine (SVM) model, which will be combined with the LSTM network to detect artifacts. Lastly, Google Colab is a cloud-based Jupyter notebook environment that allows you to write and execute Python code directly in your browser. It provides free access to GPUs and TPUs, making it a valuable resource for computationally intensive tasks like training deep learning models.

## hyes

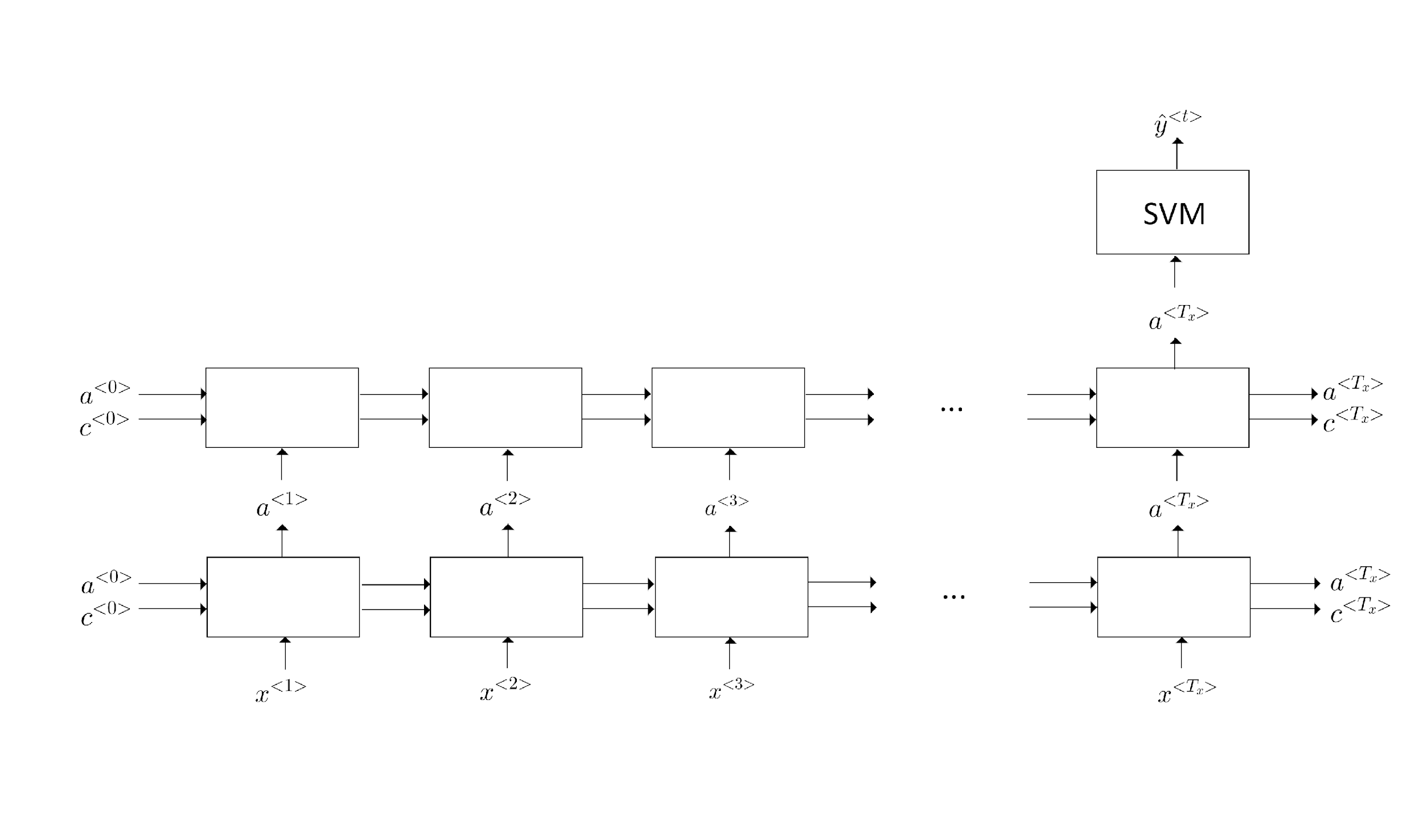
As this study will require baseline models to compare the proposed LSTM-SVM to, it will attempt to replicate the methodology used by Llanes-Jurado et al. (2023) in order to obtain baseline models that will be used to gauge the performance of the LSTM-SVM and then determine whether its performance against the latter is significant.

For Taylor et al. (2015), this study will use the replication method of Llanes-Jurado et al. (2023) of the former study for comparable results. The researchers themselves used a different dataset rather than the former. As mentioned, the researchers will use the EDABE dataset publicly made available by Llanes-Jurado et al. (2023). The signals of this dataset will be then partitioned into 0.5s segments/epochs. The EDA signal/s will also undergo feature extraction and engineering which will aim to compute the following: the minimum, maximum, mean, median, standard deviation, range of the raw signals, the minimum, maximum, mean, median, standard deviation, range of the first order derivative of the raw signals, the minimum, maximum, mean, median, standard deviation, range of the second order derivative of the raw signals. Moreover, the raw EDA signal/s will also be transformed to a low pass filtered version of 16hz, wherein the aforementioned features are then calculated again. These features can be categorized as statistical features, time frequency related features will be extracted by first applying wavelet decomposition with Haar window of level 3 (three levels) to the raw EDA signal/s. In each level the coefficients that result from the application of this decomposition is used to compute the following: the mean of the coefficients, median of the coefficients, maximum of the coefficients, standard deviation of the coefficients, and the number of coefficients above zero. Subsequently a Support Vector Machine (SVM) is then used as a backward feature selection method (BFS) to reduce potentially redundant features extracted. Using these features, the process of selecting the best model will involve three different classifiers: Logistic Regression, Support Vector Machine, and Random Forest. In order to select the best model hyperparameter with 5-fold cross validation tuning will have to be applied for each of the three classifiers e.g. 0.01 as the C hyperparameter for Logistic Regression may produce the highest validation mean AUC as opposed to using 0.1, 1, 10, and 100 during 5-fold cross validation therefore we choose the Logistic Regression model with a C hyperparameter of 0.01. Grid search algorithm is then used to find the optimal hyperparameters for each of the three classifiers. 0.01, 0.1, 1, 10 and 100 for the C hyperparameter in the Logistic Regression classifier. 200, 400, and 600 & 10, 30 and 50 for the Estimators and Max Depth hyperparameters respectively for the Random Forest classifier. 1, 10, 100 and, 1000 & 0.001, 0.01, 0.1, and 1 for the C and Gamma hyperparameters respectively in the SVM model. Once the best performing models are selected out of each of the three classifiers, the overall best performing model out of the 3 is selected based again on validation AUC. This will be selected as one of the baseline models.

Similarly, for Hossain et al. (2022), this study will use the replication method of Llanes-Jurado et al. (2023) of the former study for comparable results. The signals of the same dataset will be then partitioned into 0.5s segments/epochs. In detail the EDA signal/s will also undergo feature extraction and engineering which will aim to compute the first category of features: statistical, which would be the following: the minimum, maximum, mean, median, standard deviation, range, and Shannon entropy of the raw signals, the minimum, maximum, mean, median, standard deviation, range, and Shannon entropy of the first order derivative of the raw signals, the minimum, maximum, mean, median, standard deviation, range, and Shannon entropy of the second order derivative of the raw signals. The second category of features will involve training an autoregressive model and extract its optimized coefficients as features excluding the bias/intercept coefficient and replacing it with the error variance instead. The third category of features will involve again applying wavelet decomposition with Haar window of level 3 (three levels) to the raw EDA signal/s. In each level the coefficients that result from the application of this decomposition are used to compute the following: the mean of the coefficients, median of the coefficients, standard deviation of the coefficients, and range of the coefficients. Applying variable frequency complex demodulation (VFCDM) to the raw EDA signal/s using 64hz, 48hz, 32hz, and 16hz frequencies is also done. In each frequency band the standard deviation and mean is computed. Subsequently a Random Forest classifier is then used as a backward feature selection method (BFS) to reduce potentially redundant features extracted. Final extracted features are scaled and normalized using *StandardScaler()* and *MinMaxScaler()* (from the scikit-learn library). Using these features, the process of selecting the best model will involve four different classifiers: Logistic Regression, Gradient Boosted Tree, Random Forest and a Support Vector Machine. In order to select the best model hyperparameter with 5-fold cross validation tuning will have to be applied for each of the three classifiers. Grid search algorithm is then used to find the optimal hyperparameters for each of the three classifiers. 0.01, 0.1, 1, 10 and 100 for the C hyperparameter in the Logistic Regression classifier. 200, 400, and 600, 0.01 and 0.1 & 3, 5 and 10 for the Estimators, the Learning Rate, and the Max Depth hyperparameters respectively for the Gradient Boosted classifier. 200, 400, and 600 & 10, 30 and 50 for the Estimators and the Max Depth hyperparameters respectively for the Random Forest classifier. 1, 10, 100 and, 1000 & 0.001, 0.01, 0.1, and 1 for the C and Gamma hyperparameters respectively in the SVM model. Once the best performing models are selected out of each of the four classifiers, the overall best performing model out of the 3 is selected based again on validation AUC. This will be selected as one of the baseline models.

The last baseline model this study will compare the proposed LSTM-SVM against is the model proposed by Llanes-Jurado et al. (2023) previously that was in their study compared to the latter models by Taylor et al. (2015) and Hossain et al. (2022); the LSTM-CNN model. The signals of the same dataset will be then partitioned into 0.5s segments/epochs. These segments will then be transformed into 3D matrices the LSTM can take as input. The first two layers are LSTM layers of 16 neurons that return the hidden state output for each input time step. Subsequently, the network includes 4 convolutional levels, each of which feature 3 convolutional layers with a batch normalization operation performed after each convolution. Finally, each level included a dropout value of 0.05 and a max-pooling operation of size 2. The numbers of filters in each level were 32, 64, 128, and 256; kernel size was 5. Finally, the model features 2 fully connected layers of 256 and 16 neurons and a final fully connected layer comprising a single perceptron with a sigmoid activation function. The model will be trained with the RMSProp optimizer at a learning rate of 5 × 10−5 and a batch size of 16. Due to the imbalance of labels, the cost function that will be used to train the model will be the Dice-Sørensen coefficient (DSC). The model will have an early stopping threshold of 30 epochs. The percentage of artifacts in the training set was 12.60%. No filter or frequency band was applied to the raw signal. For each 5s segment, min–max scaling was applied.

As for the proposed approach, this study will aim to leverage the capabilities of an LSTM as a mechanism not for prediction but for automatic extraction of high level features from EDA signal data. Moreover, this study aims to integrate the simplicity and robustness of an SVM as opposed to more complex deep learning methods i.e. CNN. This approach is based on the studies by Agarap (2017) and Chen & Chen (2019) where the researchers aim to integrate the SVM mechanism in place of the typical Softmax Activation layer or a Fully Connected layer used by an RNN after it has outputted its hidden state. The signals of the same dataset will be then partitioned into 0.5s segments/epochs. These segments will then be transformed into 3D matrices the LSTM can take as input. The architecture of the LSTM-SVM is detailed as follows: the LSTM also features two layers stacked on top of the other with 16 neurons. The first LSTM layer will return the hidden state output for each input time step. Only on the last or second LSTM layer and on the last LSTM cell is a single hidden state outputted.



*Figure 14. LSTM-SVM model architecture*

Use hidden state as features to SVM mechanism model with different Gamma and C hyperparameters e.g. 1, 10, 100 and, 1000 & 0.001, 0.01, 0.1, and 1.

Evaluation will use leave-one-subject-out (LOSO) since each subject contains segments of 0.5s signals, it is imperative that the model does not see these signals of the subject that will be used for validation in order to see whether or not the model has generalized on the training dataset and can perform well on the validation dataset. We perform LOSO testing to evaluate the performance of the models, used also in studies by Di Lascio et al. (2018), Gjoreski et al. (2017), Saeb et al. (2016), Sarker et al. (2014), Schimdt et al. (2019), and Zhang et al. (2017). We keep the data of all subjects, except one, in the training set and use the remaining subject for testing the performance of the model. We repeat the same procedure for all the subjects and report the performance of the model as average score across all iterations. As there are 43 subjects, 42 will be left for training and validation. The experiment will involve splitting the training and validation dataset 70% to 30% where 30 of the subjects will be used for training and 13 will be used for validation. Using this validation procedure, shape artifact and clean segments derived from the EDA signals of a single user are not contained in the train and test simultaneously. This is a more realistic protocol because the training and test data are different due to interpersonal variance as detailed by Hernandez et al. (2011). Thereby it ensures the generalization of the model to new subjects because the models are not biased by the characteristics of particular subjects as shown in studies by Di Lascio et al. (2019) and Gjoreski et al. (2017).

## System Architecture

*Figure 15. System Architecture*

## Ethical Considerations

In conducting this research on developing a hybridized LSTM-SVM model for artifact detection, we will ensure data privacy by using anonymized, consented datasets and adhering to all licensing agreements for data usage. The model's accuracy and limitations will be transparently reported, with considerations for minimizing false positives and negatives to mitigate adverse impacts on downstream applications. We will acknowledge the dual-use nature of the technology and propose safeguards against misuse. Performance comparisons with existing models will be conducted under identical conditions using standardized metrics to ensure fairness.

## Data Analysis (Procedure and Treatment)

In order to measure and to evaluate the performance of the proposed hybridized LSTM-SVM to automatically detect the artifacts found in an EDA signal, the researchers utilized the following performance metrics:

* Area Under Curve
* Accuracy
* Precision
* F1-Score
* Recall

**Area Under Curve (AUC)**

Area Under Curve (AUC) is a proposed metric for evaluation by Llanes-Jurado et. al. (2023) to address the concern regarding the unbalanced classes in the dataset, as this metric provides a more robust measure for future comparisons utilizing the dataset. This performance metric refers to the measure of the total area covered by a curve on a graph, which allows quantification of the extent region beneath the curve between points.

**Accuracy**

Accuracy refers to how close a measurement is to the true or accepted values. In this study, it is used to evaluate the performance of the models of Taylor et. al. (2015), Hossain et. al. (2022), and Llanes-Jurado et. al. (2023), and our proposed LSTM-SVM model, in terms of the quotient of correctly predicted instances to the total number of instances.

True Positives (TP) pertains to the instances wherein artifacts are correctly detected, while True Negatives (TN) refers to the occurrence of correctly detecting non-artifacts. False Positives (FP) on the other hand pertains to incorrectly identified genuine signals as artifacts, and False Negatives (FN) are incorrectly defining artifacts as genuine EDA signals.

**Precision**

Precision refers to how close measurements of the same item are to each other. It is defined as the ratio of true positive predictions to the total number of positive predictions, which consists of both true positives and false positives. It is also utilized by the researchers of this study to assess the relevance of positive predictions, especially when the dataset is imbalanced.

**Recall**

The researchers also used recall as a performance metric to evaluate the automatic artifact detection models of Taylor et. al. (2015), Hossain et. al. (2022), and Llanes-Jurado et. al. (2023), compared to our own LSTM-SVM model’s performance. Recall can be determined with the formula:

**F1-Score**

Finally, to evaluate the given automatic artifact detection models in terms of their performance, F1-Score is also used as a metric. It is a machine learning evaluation metric that measures a model’s accuracy. It combines the precision and recall scores of a model. By using F1-score, there is a single metric that includes both characteristics of precision and recall that can be used to compare different models easily.

F1-Score

By utilizing these performance metrics, the researchers aim to address the research question of the study, which looks for the difference of the automatic artifact detection by Taylor et. al. (2015), Hossain et. al. (2022), and Llanes-Jurado et. al. (2023) in comparison to our hybridized LSTM-SVM model.

| **Artifact Detection Models** | **AUC** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- | --- | --- |
| Taylor et. al. (2015) |  |  |  |  |  |
| Hossain et. al. (2022) |  |  |  |  |  |
| Llanes-Jurado et. al. (2023) |  |  |  |  |  |
| Proposed LSTM-SMV model |  |  |  |  |  |

*Table 3. Comparison of Models in terms of AUC, Accuracy, Precision, Recall, and F1-Score*

# 

# REFERENCES

Agarap, A. F. (2018). *A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data*. https://doi.org/10.1145/3195106.3195117

Agarwal, C., Kaur, I., & Yadav, S. (2022). Hybrid CNN-SVM Model for Face Mask Detector to Protect from COVID-19. In *Lecture notes in computational vision and biomechanics* (pp. 419–426). https://doi.org/10.1007/978-981-19-0151-5\_35

Agrawal, J., Gupta, M., & Garg, H. (2021). Early Stress Detection and Analysis using EEG signals in Machine Learning Framework. *IOP Conference Series. Materials Science and Engineering*, *1116*(1), 012134. <https://doi.org/10.1088/1757-899x/1116/1/012134>

Ahuja, R., & Banga, A. (2019). Mental Stress Detection in University Students using Machine Learning Algorithms. *Procedia Computer Science*, *152*, 349–353. https://doi.org/10.1016/j.procs.2019.05.007

Albertetti, F., Simalastar, A., & Rizzotti-Kaddouri, A. (2021). Stress Detection with Deep Learning Approaches Using Physiological Signals. In *Springer eBooks* (pp. 95–111). https://doi.org/10.1007/978-3-030-69963-5\_7

Alfaras, M., Primett, W., Umair, M., Windlin, C., Karpashevich, P., Chalabianloo, N., Bowie, D., Sas, C., Sanches, P., Höök, K., Ersoy, C., & Gamboa, H. (2020). Biosensing and Actuation—Platforms coupling body Input-Output modalities for affective technologies. *Sensors*, *20*(21), 5968. https://doi.org/10.3390/s20215968

Alnazly, E., Khraisat, O. M., Al-Bashaireh, A. M., & Bryant, C. L. (2021). Anxiety, depression, stress, fear and social support during COVID-19 pandemic among Jordanian healthcare workers. *PloS One*, *16*(3), e0247679. <https://doi.org/10.1371/journal.pone.0247679>

AlShorman, O., Masadeh, M., Heyat, M. B. B., Akhtar, F., Almahasneh, H., Ashraf, G. M., & Alexiou, A. (2022). Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection. *Journal of Integrative Neuroscience*, *21*(1), 020.

<https://doi.org/10.31083/j.jin2101020>

Archana, V. R., & Devaraju, B. M. (2020, August). *View of stress detection using machine learning algorithms*. https://journal.ijresm.com/index.php/ijresm/article/view/171/154

Bach, D. R., & Friston, K. J. (2012). Model‐based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, *50*(1), 15–22. https://doi.org/10.1111/j.1469-8986.2012.01483.x

Banganho, A., Santos, M., & Da Silva, H. P. (2022). Electrodermal activity: Fundamental principles, measurement, and application. *IEEE Potentials*, *41*(5), 35–43. <https://doi.org/10.1109/mpot.2020.2983381>

Bartolomé-Tomás, A., Sánchez-Reolid, R., Fernández-Sotos, A., Latorre, J. M., & Fernández-Caballero, A. (2020). Arousal Detection in Elderly People from Electrodermal Activity Using Musical Stimuli. *Sensors*, *20*(17), 4788. https://doi.org/10.3390/s20174788

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Boucsein, W. (2012). Electrodermal activity. In *Springer eBooks*. https://doi.org/10.1007/978-1-4614-1126-0

Chen, N. W., Jaques, N., Taylor, S., Sano, A., Fedor, S., & Picard, R. W. (2015). Wavelet-based motion artifact removal for electrodermal activity. . https://doi.org/10.1109/embc.2015.7319814

Cimino, A., & Dell’Orletta, F. (2016). Tandem LSTM-SVM approach for Sentiment Analysis. In *Accademia University Press eBooks* (pp. 172–177). https://doi.org/10.4000/books.aaccademia.2003

Chen, F., & Chen, Z. D. (2019). Application of weighted combination model based on XGBoost and LSTM in sales forecasting. *Journal of Computer System Applications, 28*(10), 226–232.https://doi.org/10.15888/j.cnki.csa.007091

Chu, B., Marwaha, K., Sanvictores, T., & Ayers, D. (2019). Physiology, stress reaction. *StatPearls*. https://pubmed.ncbi.nlm.nih.gov/31082164/

Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive Assessment of Students’ Emotional Engagement during Lectures Using Electrodermal Activity Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), 1–21. https://doi.org/10.1145/3264913

Di Lascio, E., Gashi, S., & Santini, S. (2019). *Laughter Recognition Using Non-invasive Wearable Devices*. https://doi.org/10.1145/3329189.3329216

Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020, January 21). *Human Emotion*

*Recognition: Review of Sensors and Methods*. Sensors. <https://doi.org/10.3390/s20030592>

Fevre, M. L., Matheny, J., & Kolt, G. S. (2003). Eustress, distress, and interpretation in occupational stress. *Journal of Managerial Psychology*, *18*(7), 726–744. https://doi.org/10.1108/02683940310502412

Gashi, S., Di Lascio, E., Stancu, B., Swain, V. D., Mishra, V., Gjoreski, M., & Santini,

S. (2020). *Detection of Artifacts in Ambulatory Electrodermal Activity Data. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(2), 1–31.* doi:10.1145/3397316

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471. https://doi.org/10.1162/089976600300015015

Gers, F. A., Schraudolph, N., & Schmidhuber, J. (2003). *Learning Precise Timing with LSTM Recurrent Networks*. https://www.semanticscholar.org/paper/Learning-Precise-Timing-with-LSTM-Recurrent-Gers-Schraudolph/047655e733a9eed9a500afd916efa566915b9110

Giuste, F., Shi, W., Zhu, Y., Naren, T., Isgut, M., Sha, Y., Tong, L., Gupte, M., & Wang, M. D. (2023). Explainable Artificial Intelligence Methods in Combating Pandemics: A Systematic review. *IEEE Reviews in Biomedical Engineering*, *16*, 5–21. https://doi.org/10.1109/rbme.2022.3185953

Gjoreski, M., Luštrek, M., Gams, M., & Gjoreski, H. (2017). Monitoring stress with a wrist device using context. *Journal of Biomedical Informatics*, *73*, 159–170. https://doi.org/10.1016/j.jbi.2017.08.006

Greco, A., Valenza, G., Lanata, A., Scilingo, E., & Citi, L. (2016). cvxEDA: a Convex Optimization Approach to Electrodermal Activity Processing. *IEEE Transactions on Bio-medical Engineering/IEEE Transactions on Biomedical Engineering*, 1. https://doi.org/10.1109/tbme.2015.2474131

Guo, X., Yin, Y., Dong, C., Yang, G., & Zhou, G. (2008). On the Class Imbalance Problem. *2008 Fourth International Conference on Natural Computation, Jinan, China*. https://doi.org/10.1109/icnc.2008.871

Hemakom, A., Atiwiwat, D., & Israsena, P. (2023). ECG and EEG based detection and multilevel classification of stress using machine learning for specified genders: A preliminary study. *PloS One*, *18*(9), e0291070. https://doi.org/10.1371/journal.pone.0291070

Hernandez, J., Morris, R. R., & Picard, R. W. (2011). Call Center Stress Recognition with Person-Specific Models. In *Lecture notes in computer science* (pp. 125–134). https://doi.org/10.1007/978-3-642-24600-5\_16

Heyat, M. B. B., Akhtar, F., Abbas, S. J., Al-Sarem, M., Alqarafi, A., Stalin, A., Abbasi, R., Muaad, A. Y., Lai, D., & Wu, K. (2022). Wearable flexible electronics based cardiac electrode for researcher mental stress detection system using machine learning models on single lead electrocardiogram signal. *Biosensors*, *12*(6), 427. <https://doi.org/10.3390/bios12060427>

Hernandez, J., Morris, R. R., & Picard, R. W. (2011). Call Center Stress Recognition with Person-Specific Models. In *Lecture notes in computer science* (pp. 125–134). https://doi.org/10.1007/978-3-642-24600-5\_16

Hicks, S., Riegler, M., Pogorelov, K., Anonsen, K. V., De Lange, T., Johansen, D., Jeppsson, M., Randel, K. R., Eskeland, S. L., & Halvorsen, P. (2018). *Dissecting Deep Neural Networks for Better Medical Image Classification and Classification Understanding*. https://doi.org/10.1109/cbms.2018.00070

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hossain, M. B., Posada-Quintero, H. F., Kong, Y., McNaboe, R., & Chon, K. H. (2022, April 1). *Automatic motion artifact detection in electrodermal activity data using machine learning*. Biomedical Signal Processing and Control. https://doi.org/10.1016/j.bspc.2022.103483

Hussein, R., Palangi, H., Ward, R. K., & Wang, Z. J. (2019). Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. *Clinical Neurophysiology*, *130*(1), 25–37. <https://doi.org/10.1016/j.clinph.2018.10.010>

Jamshidzadeh, Z., Ehteram, M., & Shabanian, H. (2024). Bidirectional Long Short-Term Memory (BILSTM) - Support Vector Machine: A new machine learning model for predicting water quality parameters. *Ain Shams Engineering Journal/Ain Shams Engineering Journal*, *15*(3), 102510. https://doi.org/10.1016/j.asej.2023.102510

Jaseena, K., & Kovoor, B. C. (2021). Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. *Energy Conversion and Management*, *234*, 113944. https://doi.org/10.1016/j.enconman.2021.113944

Jose, J. (2022). INTRODUCTION TO TIME SERIES ANALYSIS AND ITS APPLICATIONS.*ResearchGate*.https://www.researchgate.net/publication/362389180\_INTRODUCTION\_TO\_TIME\_SERIES\_ANALYSIS\_AND\_ITS\_APPLICATIONS

Joseph, R. V., Mohanty, A., Tyagi, S., Mishra, S., Satapathy, S. K., & Mohanty, S. N. (2022). A hybrid deep learning framework with CNN and Bi-directional LSTM for store item demand forecasting. *Computers & Electrical Engineering*, *103*, 108358. https://doi.org/10.1016/j.compeleceng.2022.108358

Juneau, P., Baddour, N., Burger, H., Bavec, A., & Lemaire, E. D. (2021). Comparison of Decision Tree and Long Short-Term memory approaches for automated foot strike detection in lower extremity amputee populations. *Sensors*, *21*(21), 6974. https://doi.org/10.3390/s21216974

Khorram, A., Khalooei, M., & Rezghi, M. (2020). End-to-end CNN + LSTM deep learning approach for bearing fault diagnosis. *Applied Intelligence*, *51*(2), 736–751. https://doi.org/10.1007/s10489-020-01859-1

Kleckner, I. R., Jones, R. M., Wilder-Smith, O., Wormwood, J. B., Akcakaya, M., Quigley, K. S., Lord, C., & Goodwin, M. S. (2018). Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data. *IEEE Transactions on Bio-medical Engineering/IEEE Transactions on Biomedical Engineering*, *65*(7), 1460–1467. https://doi.org/10.1109/tbme.2017.2758643

Koklu, M., Unlersen, M. F., Ozkan, I. A., Aslan, M. F., & Sabanci, K. (2022). A CNN-SVM study based on selected deep features for grapevine leaves classification. *Measurement*, *188*, 110425. https://doi.org/10.1016/j.measurement.2021.110425

Kortum, E., Leka, S., & Cox, T. (2010). Psychosocial risks and work-related stress in developing countries: health impact, priorities, barriers and solutions. *International Journal of Occupational Medicine and Environmental Health*, *23*(3). https://doi.org/10.2478/v10001-010-0024-5

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). International Affective Picture System (IAPS) Affective Ratings of Pictures and Instruction Manual. Technical Report A-6, Gainesville, FL University of Florida. - References - Scientific Research Publishing. (n.d.). https://www.scirp.org/reference/referencespapers?referenceid=2323896

Lee, G., Choi, B., Jebelli, H., Ahn, C. R., & Lee, S. (2020). Noise reference Signal–Based denoising method for EDA collected by multimodal biosensor wearable in the field. *Journal of Computing in Civil Engineering*, *34*(6). https://doi.org/10.1061/(asce)cp.1943-5487.0000927

Li, S., Zhao, X., Stankovic, L., & Mandic, D. (2022). Demystifying CNNs for images by matched filters. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2210.08521

Lin, X., Lei, Y., Chen, J., Xing, Z., Yang, T., Wang, Q., & Wang, C. (2023). A Case-Finding Clinical Decision Support System to Identify Subjects with Chronic Obstructive Pulmonary Disease Based on Public Health Data. *Tsinghua Science and Technology/Tsinghua Science and Technology*, *28*(3), 525–540. https://doi.org/10.26599/tst.2022.9010010

Liu, Y., & Du, S. (2018). Psychological stress level detection based on electrodermal activity. *Behavioural Brain Research*, *341*, 50–53. <https://doi.org/10.1016/j.bbr.2017.12.021>

Llanes-Jurado, J., Carrasco-Ribelles, L. A., Alcañiz, M., & Marín-Morales, J. (2021). Automatic artifact recognition and correction for electrodermal activity in uncontrolled environments. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-717360/v

Llanes-Jurado, J., Carrasco-Ribelles, L. A., Alcañiz, M., Soria-Olivas, E., & Marín-Morales, J. (2023, November 1). *Automatic artifact recognition and correction for electrodermal activity based on LSTM-CNN models*. Expert Systems With Applications. https://doi.org/10.1016/j.eswa.2023.120581

Mordensky, S. P., Lipor, J. J., DeAngelo, J., Burns, E. R., & Lindsey, C. R. (2023). When less is more: How increasing the complexity of machine learning strategies for geothermal energy assessments may not lead toward better estimates. *Geothermics*, *110*, 102662. <https://doi.org/10.1016/j.geothermics.2023.102662>

Moruzzi, G., & Magoun, H. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, *1*(1–4), 455–473. https://doi.org/10.1016/0013-4694(49)90219-9

Mohsen, S., Ghoneim, S. S. M., Alzaidi, M. S., Alzahrani, A., & Hassan, A. M. A. (2023). Classification of electroencephalogram signals using LSTM and SVM based on Fast Walsh-Hadamard Transform. *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, *75*(3), 5271–5286. https://doi.org/10.32604/cmc.2023.038758

Mozer, M. C. (1997). Induction of multiscale temporal structure. *ResearchGate*. https://www.researchgate.net/publication/2926596\_Induction\_of\_Multiscale\_Temporal\_Structure

Munir, M., Siddiqui, S. A., Küsters, F., Mercier, D., Dengel, A., & Ahmed, S. (2019). TSXPlain: Demystification of DNN decisions for Time-Series using natural language and statistical features. In *Lecture notes in computer science* (pp. 426–439). https://doi.org/10.1007/978-3-030-30493-5\_43

Mordensky, S. P., Lipor, J. J., DeAngelo, J., Burns, E. R., & Lindsey, C. R. (2023). When less is more: How increasing the complexity of machine learning strategies for geothermal energy assessments may not lead toward better estimates. *Geothermics*, *110*, 102662. <https://doi.org/10.1016/j.geothermics.2023.102662>

Moruzzi, G., & Magoun, H. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, *1*(1–4), 455–473. https://doi.org/10.1016/0013-4694(49)90219-9

Nagabushanam, P., George, S. T., & Radha, S. (2019). EEG signal classification using LSTM and improved neural network algorithms. *Soft Computing*, *24*(13), 9981–10003. https://doi.org/10.1007/s00500-019-04515-0

Orlanda-Ventayen, C., & Ventayen, R. J. (2022). Stress and depression in the workplace of educators in the Philippines. *International Journal of Occupational Safety and Health*, *12*(4), 325–335. https://doi.org/10.3126/ijosh.v12i4.43226

O’Reilly, R. C., & Frank, M. J. (2006). Making working Memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328. https://doi.org/10.1162/089976606775093909

Park, Y., & Yang, H. S. (2019). Convolutional neural network based on an extreme learning machine for image classification. *Neurocomputing*, *339*, 66–76. https://doi.org/10.1016/j.neucom.2018.12.080

Pérez-Ortiz, J. A., Gers, F. A., Eck, D., & Schmidhuber, J. (2003). Kalman filters improve LSTM network performance in problems unsolvable by traditional recurrent nets. *Neural Networks*, *16*(2), 241–250. https://doi.org/10.1016/s0893-6080(02)00219-8

Posada-Quintero, H. F., & Chon, K. H. (2020). Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic review. *Sensors*, *20*(2), 479. https://doi.org/10.3390/s20020479

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(03). https://doi.org/10.1017/s0954579405050340

Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., & Kording, K. P. (2016). Voodoo machine learning for clinical predictions. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/059774

Sarker, H., Sharmin, M., Ali, A. A., Rahman, M. M., Bari, R., Hossain, S. M., & Kumar, S. (2014). *Assessing the availability of users to engage in just-in-time intervention in the natural environment*. https://doi.org/10.1145/2632048.2636082

Sánchez-Reolid, R., De La Rosa, F. L., Sánchez-Reolid, D., López, M. T., & Fernández-Caballero, A. (2022). Machine Learning Techniques for Arousal Classification from Electrodermal Activity: A Systematic Review. *Sensors*, *22*(22), 8886. https://doi.org/10.3390/s22228886

Sánchez-Reolid, R., Martínez-Rodrigo, A., & Fernández-Caballero, A. (2019). Stress Identification from Electrodermal Activity by Support Vector Machines. In *Lecture notes in computer science* (pp. 202–211). https://doi.org/10.1007/978-3-030-19591-5\_21

Sánchez-Reolid, R., Martínez-Rodrigo, A., López, M. T., & Fernández-Caballero, A. (2020). Deep Support Vector Machines for the Identification of Stress Condition from Electrodermal Activity. *International Journal of Neural Systems*, *30*(07), 2050031. <https://doi.org/10.1142/s0129065720500318>

Semmlow, J. L. (2004). *Biosignal and medical image processing*. CRC Press.

Schmidt, P., Dürichen, R., Reiss, A., Van Laerhoven, K., & Plötz, T. (2019). Multi-target affect detection in the wild:An Exploratory Study. *Conference: The 23rd International Symposium*. <https://doi.org/10.1145/3341163.3347741> *Stress*. (2022, June 17). https://www.who.int/news-room/questions-and-answers/item/stress#:~:text=Stress%20can%20be%20defined%20as,experiences%20stress%20to%20some%20degree

Sun, L., Jin, B., Yang, H., Tong, J., Liu, C., & Xiong, H. (2018). Unsupervised EEG feature extraction based on echo state network. *Information Sciences*, *475*, 1–17. https://doi.org/10.1016/j.ins.2018.09.057

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). *Rethinking the Inception Architecture for Computer Vision*. https://doi.org/10.1109/cvpr.2016.308

Tang, Y. (2013). Deep Learning using Linear Support Vector Machines. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1306.0239>

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. Annu Int Conf IEEE Eng Med Biol Soc. 2015;2015:1934-7. doi: 10.1109/EMBC.2015.7318762. PMID: 26736662; PMCID: PMC5413200.

Thammasan, N., Stuldreher, I. V., Schreuders, E., Giletta, M., & Brouwer, A. (2020). A usability study of physiological measurement in school using wearable sensors. *Sensors*, *20*(18), 5380. https://doi.org/10.3390/s20185380

Vapnik, V. N. (1995). The nature of statistical learning theory. In *Springer eBooks*. https://doi.org/10.1007/978-1-4757-2440-0

Wan, H., Xu, R., Zhang, M., Cai, Y., Li, J., & Shen, X. (2022). A novel model for water quality prediction caused by non-point sources pollution based on deep learning and feature extraction methods. *Journal of Hydrology*, *612*, 128081. https://doi.org/10.1016/j.jhydrol.2022.128081

Wang, B., Ma, R., Kuang, J., & Zhang, Y. (2020). How decisions are made in brains: Unpack “Black Box” of CNN with Ms. Pac-Man video game. *IEEE Access*, *8*, 142446–142458. <https://doi.org/10.1109/access.2020.3013645>

Wang, Q.S.. (2019). Research and application of timing optimization of short-term traffic flow at urban intersections. University of Science and Technology of China, Hefei, China.

Weerakody, P. B., Wong, K. W., Wang, G., & Ela, W. (2021). A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*, *441*, 161–178. https://doi.org/10.1016/j.neucom.2021.02.046

Wu, Z., Rincon, D., & Christofides, P. D. (2020). Process structure-based recurrent neural network modeling for model predictive control of nonlinear processes. *Journal of Process Control*, *89*, 74–84. https://doi.org/10.1016/j.jprocont.2020.03.013

Yanase, J., & Triantaphyllou, E. (2019). The seven key challenges for the future of computer-aided diagnosis in medicine. *International Journal of Medical Informatics*, *129*, 413–422. https://doi.org/10.1016/j.ijmedinf.2019.06.017

Zhang, Q., Wang, R., Qi, Y., & Wen, F. (2022). A watershed water quality prediction model based on attention mechanism and Bi-LSTM. *Environmental Science and Pollution Research International*, *29*(50), 75664–75680. https://doi.org/10.1007/s11356-022-21115-y

Zhang, Y., Haghdan, M., & Xu, K. S. (2017). Unsupervised motion artifact detection in Wrist-Measured Electrodermal Activity data. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1707.08287

# 

# APPENDIX

**System Mockup**

**Main Page** 