Stress Detection Using CNN on the WESAD Dataset

D. Shirley Benita
Department of Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
dshirley@karunya.edu.in

A. Shamila Ebenezer*
Department of Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
shamila_cse@karunya.edu
(*Communicating Author)*

Susmitha L
Department of Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
lsusmitha94@gmail.com

M.S.P.Subathra
Departnemt of Robotics Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
subathra@karunya.edu

S. Jeba Priya
Department of Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, India
jebapriya@karunya.edu

2024 International Conference on Emerging Systems and Intelligent Computing (ESIC) | 979-8-3503-4985-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ESIC60604.2024.10481604

Abstract*- Stress, a pervasive psychological state, profoundly impacts individuals' well-being and may lead to various health issues. In this paper, the WESAD (Wearable Stress and Affect Detection) dataset serves as the primary data source, from which HRV (Heart Rate Variability) data extracted using PPG (Photoplethysmography) sensor is utilized. This data can be obtained in a non-invasive method and is instrumental in this analysis. The dataset comprises HRV data from 15 individual test subjects obtained from PPG sensors, providing a crucial foundation for the detection and classification of individuals into "stressed" or "not stressed" categories, thus offering a valuable tool for stress monitoring. This paper is a focused exploration into the CNN (Convolutional Neural Network) Deep Learning model, in which the feature selection process is carried out using Random Forest resulting in the selection of HR (Heart Rate), LF_BOXCOX (Low-Frequency Component - Box-Cox Transformed), MEDIAN_RR (Median of RR Intervals), SD1_BOXCOX (Standard Deviation 1 - Box-Cox Transformed), and SDRR_REL_RR (Standard Deviation of Relative RR Intervals) to optimize the model's performance. The model aims at advancing and refining stress detection methodologies to enhance overall well-being and achieves an accuracy of 98%.

Keywords: Stress, WESAD Dataset, PPG sensor, non-invasive method, HRV, CNN, feature selection, Random Forest, HR, LF_BOXCOX, MEDIAN_RR, SD1_BOXCOX, and SDRR_REL_RR.

I. INTRODUCTION

Stress is an omnipresent force that we encounter in our daily lives, intricately woven into the fabric of our activities and experiences. It is a dynamic and often inevitable aspect of the human condition, capable of influencing our physical and emotional well-being in profound ways. Acknowledging and understanding the role of stress in our lives is pivotal, as it offers a window into the depths of our psyche and can be a key to unlocking our true potential. By recognizing and effectively managing stress, we gain valuable insights into ourselves, which, when harnessed skillfully, can empower us to navigate life's challenges with greater resilience, clarity, and advantage.

The identification of stress is a formidable challenge, primarily owing to its inherently subjective and multifaceted nature. Traditional methods for detecting stress have often relied on intrusive and sometimes obtrusive approaches, such as questionnaires, biochemical markers, and psychological evaluations. These methods not only disrupt the natural flow of a person's daily life but also introduce elements of subjectivity and bias that can limit their effectiveness. In contrast, the advent of non-intrusive technologies offers a promising avenue for stress detection. Among these, the measurement of HRV (Heart Rate Variability) has emerged as potent tools. PPG (Photoplethysmography) sensors, which can be integrated into wearable devices like smartwatches, provide a non-invasive means of monitoring changes in blood flow, which are closely linked to stress responses. HRV offers a nuanced insight into an individual's autonomic nervous system, providing a more accurate and objective measure of stress levels.

In this paper, HRV values sourced from the WESAD dataset are considered for research. The dataset is then used in the investigation of stress detection through Deep learning (DL) methodologies. At the outset, the primary objective of this paper was to employ a foundational 1D CNN (Convolutional Neural Network) model, which entailed the processing of all input features through the said model. This approach yielded negligible utility, as the model exhibited a pronounced inability to acquire meaningful insights from the data. Consequently, a revised strategy was adopted, wherein the CNN model exclusively processed a subset of features identified by the Random Forest classifier, a technique that significantly facilitated the discernment of latent patterns within the dataset. Following a meticulous process of hyperparameter optimization, this refined model achieved a notable accuracy rate of 98%. This demonstrates the capacity of DL methods to harness the intricate patterns within HRV data, providing a more robust and accurate framework for stress detection.

A. Related Works

In [1], the authors present the SWELL-KW dataset, which acts as a resource for research on stress and user modeling. The dataset collection was performed through an experiment consisting of 25 individuals engaged in knowledge work tasks under controlled stress conditions, including email interruptions and time pressure. It encompasses a wide range of data sources, such as computer logging, facial expressions, body postures, heart rate variability, and skin conductance. Questionnaires were provided to gauge the subjects' subjective experiences related to task load, mental effort, emotion, and perceived stress. This comprehensive dataset contributes significantly to work psychology, user modeling, and context-aware systems by offering insights into the complex interplay between stress and knowledge work.

In [2], the research explores the pervasive issue of workplace stress and the potential of context-aware systems to alleviate stress among knowledge workers. The central focus is on the development of automatic classifiers that can identify both working and stress-related mental state by utilizing a multimodal sensor dataset, encompassing computer logging, facial expressions, posture, and physiology. This investigation addresses two critical challenges in machine learning: the detection of work-related stress using minimally intrusive sensors and the consideration of individual differences. Notably, the findings reveal that support vector machines (SVM) can distinguish neutral and stressful conditions with 90% accuracy on the SWELL-KW dataset, with posture providing the most informative data, followed closely by facial expressions. Furthermore, the paper sheds light on the ability of sensor data to predict 'mental effort' more effectively than 'perceived stress,' achieving the highest correlation through a decision tree method. This research underscores the potential of advanced sensor technologies and machine learning to yield valuable insights into stress detection and the nuances of mental states in the workplace, offering promising avenues for developing tailored support systems for knowledge workers.

In [3], the authors introduce the WESAD dataset, designed to advance affect recognition and wearable stress detection for human-computer interaction and well-being. This dataset includes physiological and motion data from wrist- and chest-worn devices, featuring various sensor modalities. Notably, it encompasses three affective states—neutral, stress, and amusement—and incorporates self-reported subject assessments. The paper establishes a classification benchmark, by achieving an accuracy of about 80% for the three-class problem and 93% for the binary stress versus non-stress classification. The study also provides a detailed analysis of device locations and sensor modalities, offering significant contributions to the field of affective computing and stress monitoring.

In [4], the researchers implement a self-supervised DL approach to enhance ECG based emotion recognition. The method involves two learning stages: first, it learns abstract ECG representations through a network that recognizes signal transformations, and second, it focuses on classifying emotions. The network gains spatiotemporal representations by performing pretext tasks involving six different signal transformations on unlabeled ECG data. These learned representations are then passed on to an emotion recognition network, where the CNN layers remain fixed, and the dense layers are trained by utilizing labeled ECG data. Notably, this approach demonstrates significant performance improvements compared to fully supervised learning, achieving tremendous results in various emotion categories across multiple datasets. The study offers insights into the effectiveness of this structure and the optimal difficulty levels for pretext self-supervised tasks, which contribute in the field of emotion recognition.

In [5], the authors investigate the issues related to stress and the importance of detecting stress in its initial stages. To address this, the paper proposes to employ machine learning and deep learning for stress detection using data from wearable sensors. The data used in this study is the WESAD dataset. This paper employs classification algorithms like SVM, DT, and KNN, along with deep learning models such as Long Short-Term Memory (LSTM). The evaluation of these models is performed based on the confusion metrics parameters, including accuracy, precision, recall, and f1-score. This research offers a promising approach to help individuals mitigate the adverse health effects of stress by providing early detection and intervention using wearable technology.

In [6], the paper implores the two approaches to training stress detection models: subject-dependent and subject-independent methods. While subject-dependent models offer the highest accuracy, subject-independent models are more practical, allowing the implementation of stress detection systems in everyday wearable devices without the need for user-specific training data. To improve subject-independent stress detection, bio-signal processing pipeline and a simple neural network using statistical features from multiple sensing sources, including Electrodermal Activity (EDA), Blood Volume Pulse (BVP), and Skin Temperature (ST) is introduced in the paper. Then evaluation of the accuracy of these models that utilize data from individual signal sources and those that fuse data from multiple sources is performed. Experiments conducted with the WESAD dataset demonstrate that the proposed model outperforms existing methods, achieving a 1.63% higher mean accuracy compared to the leading model, while maintaining a low standard deviation. Furthermore, the study highlights that combining features from multiple sources results in more accurate predictions compared to using individual sensor sources. This research contributes to the advancement of subject-independent stress detection methods, particularly in the context of wearable devices and their potential to enhance human well-being.

B. *Objective*

The main objectives of this project are given as:

1. Develop a stress detection system with the objective of simplifying the complex problem into a binary classification, enabling easy identification of individuals as "stressed" or "not stressed."
2. Investigate the application of CNN in the realm of stress detection, specifically focusing on binary classification to enhance the accuracy and efficiency of stress identification.
3. Design and implement a user-friendly stress monitoring tool that utilizes the CNN model, ensuring fast and reliable results for individuals. The aim is to provide users with actionable insights to improve their mental state and overall well-being based on the detected stress levels.

II. METHODOLOGY

As shown in Fig 1, this paper focuses on utilizing the WESAD dataset as a primary resource in determining whether an individual is experiencing stress or not. It primarily revolves around commencing with a fundamental 1D CNN model and subsequently enhancing its performance through the implementation of advanced methodologies such as feature selection and meticulous hyperparameter tuning.

The paper initiates by engaging in the preprocessing of the data derived from the WESAD dataset. Subsequent to the meticulous preprocessing and refinement of the dataset, it is partitioned into training and testing subsets, observing a ratio of 3:1. The training data is subsequently employed as input for the 1D CNN model, while the testing data is utilized to procure and evaluate the outcomes.

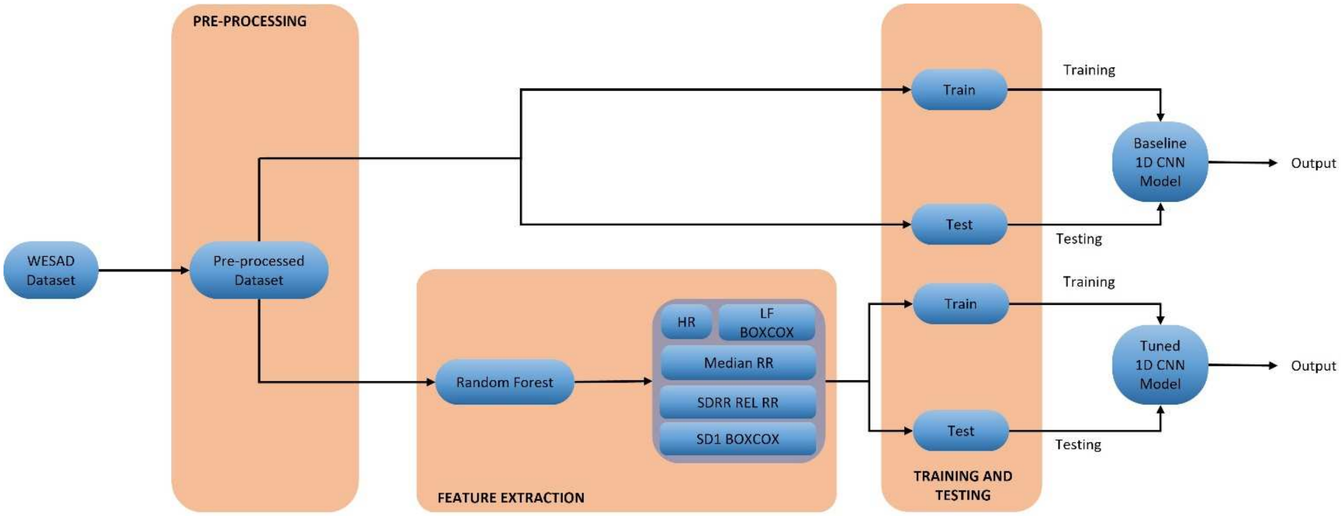

Fig. 1. Workflow for Binary Stress Classification using 1D CNN

A. *Dataset Description*

This paper utilizes the WESAD dataset from [3]. The WESAD dataset, is a publicly available multivariate dataset that comprises physiological and motion sensor data collected from 15 test subjects. The dataset comprises of various sensor information, such as blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration. Additionally, it classifies the subjects into three states: neutral, stressed, and amused.

In this paper, the data acquired from the 15 individual test subjects has been aggregated and collectively subjected to the subsequent processing and analysis procedures. The HRV-related segment of the WESAD dataset obtained from PPG sensor comprises a total of 1,35,650 records and 67 columns. Within this dataset, two columns contain object data, while the remaining 65 columns consist of floating-point data.

As part of the pre-processing phase, the dataset is streamlined to encompass 112,586 records and 66 columns. To address data imbalance, the majority class is downsampled, resulting in an equally balanced dataset containing 40,946 records for both classes. Subsequently, the dataset is further refined to comprise 81,892 records and 65 columns to align with the model's requirements and necessitated pre-processing steps.

B. *Preprocessing*

Within the pre-processing phase of this work, several tasks were performed to prepare the dataset for subsequent analysis. Firstly, the unwanted records corresponding to the "amusement condition" (label 1) are systematically dropped from the dataset as this paper mainly focuses on the binary classification problem. The remaining labels were converted within a binary range. This selective elimination serves to streamline the dataset, removing data points that are not relevant to the specific focus of stress detection.

Another substantial challenge addressed during pre-processing is the presence of imbalanced data within the dataset. To tackle this issue, a strategic approach is employed to balance the dataset. This involves equalizing the number of records for both classes present in the data – in this case, the "stressed" and "not stressed" categories. To achieve this balance, some instances of the majority class are randomly dropped. This balancing act ensures that the machine learning models applied in subsequent analyses can perform optimally and make accurate predictions, despite the inherent data imbalance, thereby enhancing the reliability and robustness of the results generated in this study.

C. *Feature Selection*

In the context of this paper, when addressing the baseline model, the entire dataset is directly input into the model without any reduction in the number of features.

Upon concluding the working of the baseline model, subsequent steps were undertaken to enhance its performance. Instead of sending the complete dataset directly, an initial feature selection process is undertaken to streamline the data and enhance the efficiency of the CNN model. To accomplish this, a Random Forest model is employed on the various columns within the dataset. The primary objective here is to identify and extract the most relevant and informative features from the extensive array of available data points. The Random Forest model employs the Gini index to discern the prominent features, aligning with the paper's focus on a classification problem.

The selected subset of features, extracted from the original WESAD dataset, comprises HR (Heart Rate), LF_BOXCOX (Low-Frequency Component - Box-Cox Transformed), MEDIAN_RR (Median of RR Intervals), SD1_BOXCOX (Standard Deviation 1 - Box-Cox Transformed), and SDRR_REL_RR (Standard Deviation of Relative RR Intervals). These meticulously chosen features have been meticulously curated to serve as the refined input for our Deep Learning (DL) model.

This strategic feature selection process results in a more streamlined and focused feature set, optimizing the DL model's capabilities. It facilitates a nuanced analysis of the stress-related factors we are investigating, enhancing our ability to discern and comprehend these intricate aspects of our research.

D. Algorithms Used

Random Forest- It operates by creating an ensemble of decision trees, each trained on a subset of the data and features. In feature selection, Random Forest can be a powerful tool. It assigns an importance score to each feature by measuring how much they contribute to the model's accuracy. Features possessing higher importance scores are considered more critical, while those with lower scores can be candidates for removal.

By selecting the top-ranked features, Random Forest aids in simplifying models, enhancing interpretability, and potentially boosting predictive performance. In this paper, the model uses the Gini Index/ Entropy as used for classification problems. The formulas for the same are given in Equation (1) and (2).

$$I_G = 1 - \sum_{i=1}^{c} (p_i)^2 \quad (1)$$

$$I_H = - \sum_{i=1}^{c} p_i log_2 p_i \quad (2)$$

1D CNN- Convolutional Neural Network is a structured form of feed-forward neural network, which autonomously acquires feature engineering through the optimization of filters. The model in this study operates within a supervised learning framework. The 1D model consists of filters and pooling layers that analyze data along a single axis. The primary role of the CNN in this research is to pinpoint the specific regions of interest within the dataset.

The tuned model is provided with carefully extracted features that guide its focus towards these specific regions, allowing it to uncover hidden relationships and patterns related to the features of interest. The model structure involves a series of 1D Convolutional layers followed by MaxPooling layers, culminating in a Flatten layer and a Dense layer, ultimately leading to the Output layer. This configuration facilitates efficient data analysis, particularly for accurate stress detection.

TABLE I. BASELINE 1D CNN ARCHITECTURE

Layer	*Type*	*Input Shape*	*Output Shape*	*No. of Kernel*	*Kernel size*	*Stride*	*Padding*	*Activation*
1	Convolutional	(64, 1)	(62, 64)	64	3	1	Valid	ReLu
2	Pooling	(62, 64)	(31, 64)	None	2	None	Valid	None
3	Convolutional	(31, 64)	(29, 128)	128	3	1	Valid	ReLu
4	Pooling	(29, 128)	(14, 128)	None	2	None	Valid	None
5	Convolutional	(14, 128)	(12, 256)	256	3	1	Valid	ReLu
6	Pooling	(12, 256)	(6, 256)	None	2	None	Valid	None
7	Convolutional	(6, 256)	(4, 128)	128	3	1	Valid	ReLu
8	Pooling	(4, 128)	(2, 128)	None	2	None	Valid	None
9	Flatten	(2, 128)	(256)	None	None	None	None	None
10	Dense	(256)	(32)	None	None	None	None	ReLu
11	Dense	(32)	(1)	None	None	None	None	Sigmoid

TABLE II. TUNED 1D CNN ARCHITECTURE

Layer	*Type*	*Input Shape*	*Output Shape*	*No. of Kernel*	*Kernel size*	*Stride*	*Padding*	*Activation*
1	Convolutional	(5, 1)	(5, 32)	32	1	1	Valid	ReLu
2	Pooling	(5, 32)	(5, 32)	None	1	None	Valid	None
3	Convolutional	(5, 32)	(5, 64)	64	1	1	Valid	ReLu
4	Pooling	(5, 64)	(5, 64)	None	1	None	Valid	None
5	Convolutional	(5, 64)	(5, 128)	128	1	1	Valid	ReLu
6	Pooling	(5, 128)	(5, 128)	None	1	None	Valid	None
7	Convolutional	(5, 128)	(5, 64)	64	1	1	Valid	ReLu
8	Pooling	(5, 64)	(2, 64)	None	2	None	Valid	None
9	Flatten	(2, 64)	(128)	None	None	None	None	None
10	Dense	(128)	(32)	None	None	None	None	ReLu
11	Dense	(32)	(1)	None	None	None	None	Sigmoid

III. IMPLEMENTATION

A. Training and Testing

The initial CNN model (provided in Table I) operates on the entire dataset, encompassing a set of 66 meticulously extracted features, which serve as the model's input. This foundational CNN architecture comprises four successive iterations of 1D Convolutional layers, each followed by 1D MaxPooling layers. These layers employ filter configurations of 64, 128, 256, and 128, respectively. Subsequently, the data flows into a Flatten layer, seamlessly transitioning into a Dense layer, culminating in the binary Output layer.

In contrast, as stated in Table II, the fine-tuned CNN model only works on the subset of features obtained from feature selection and adheres to the same sequence of 1D Convolutional and 1D MaxPooling layers but employs refined kernel configurations of 32, 64, 128, and 64. After these layers, the data is directed into a Flatten layer, where the features are unwrapped and passed into a Dense layer. This Dense layer incorporates the rectified linear unit (ReLU) activation function and is succeeded by a sigmoid Output Layer.

Both iterations of the CNN models employ the binary cross-entropy loss function and the Adam optimizer, providing a robust framework for the intended tasks.

The baseline model undergoes 30 epochs, while the tuned model is subjected to an extended training duration of 100 epochs. Notably, a validation split of 0.2 is applied, and a batch size of 32 is utilized during the training process of both models. The performance metrics employed in testing of the models are provided below in Equations (3)-(7):

Accuracy= (TN + TP) / (TN+TP+FN+FP) (3)

Precision= TP / (TP + FP) (4)

Recall= TP / (TP + FN) (5)

f1 score= 2[(precision*recall)/(precision+ recall)]* (6)

Binary-cross entropy Loss = $-\frac{1}{N}\sum_i^N \sum_j^M y_{ij} \log(p_{ij})$ (7)

The baseline model achieves an accuracy of 49.9 during training with a loss of 0.6932. The parameters are thoughtfully fine-tuned to optimize the model's accuracy. As a result, the tuned model achieves a remarkable training accuracy of 98.22% with a corresponding loss of 0.0454.

The table (Table III) given below provides an overview of the performance of the models during testing on the given metrics:

Table III. PERFORMANCE METRICS OF MODELS

Model	Accuracy	Precision	Recall	F1	Loss
Baseline	49.83	0	nan	nan	0.693
Tuned	98.55	41.41	1	58.57	0.0391

Fig 2 shows the Training and Loss Curves for these 1D CNN models.

.

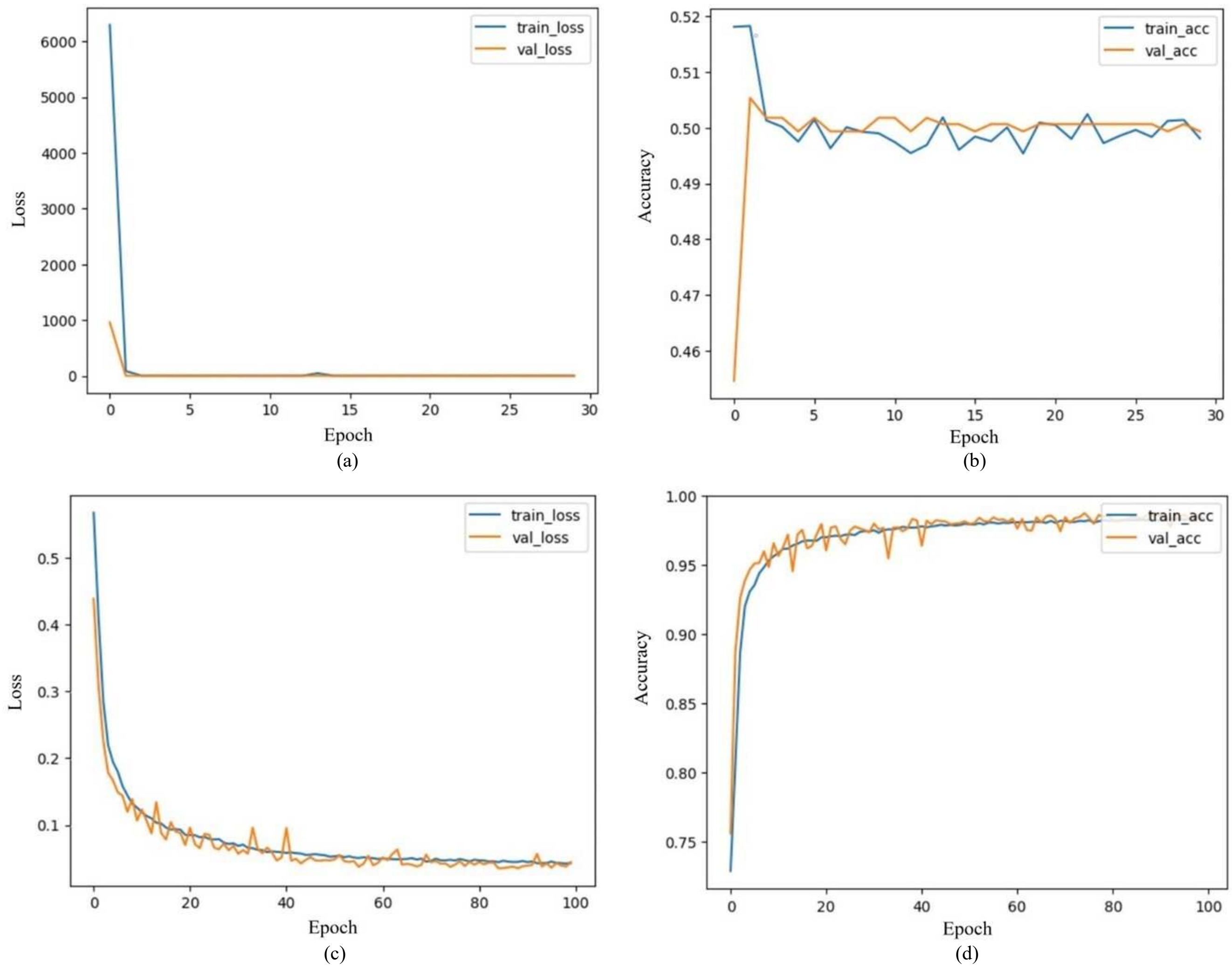

Fig. 2. a) Traning vs Validation Loss for Baseline Model f) Training vs Validation Accuracy for Baseline Model c) Traning vs Validation Loss for Tuned Model d) Training vs Validation Accuracy for Tuned Model

IV. Conclusion

This study has successfully implemented the 1D CNN Deep Learning model using the WESAD dataset with a primary focus on stress detection, primarily relying on HRV data obtained from PPG sensors in this dataset. This research carries significant implications for the detection of stress in individuals, thereby contributing to their understanding of mental health and overall well-being.

Distinguishing itself from traditional methodologies that employ intrusive instruments to collect physiological data from test subjects, this approach prioritizes the use of non-intrusive sensors. It harnesses data from these sensors to make accurate predictions and inferences. Particularly noteworthy is the effectiveness of the tuned 1D CNN model in this context, which attains an impressive overall accuracy of 98.26% utilizing features such as HR, LF_BOXCOX, MEDIAN_RR, SD1_BOXCOX, and SDRR_REL_RR obtained using Random Forest feature selection. The refined model is distinguished by a notable reduction in the count of trainable parameters in contrast to the baseline model. Moreover, it manifests a superior accuracy when compared with the baseline counterpart.

The discernible efficacy of the tuned model in mitigating the misclassification issue is apparent. Further refinement to address this concern can be achieved through an expansion of the dataset size. The dataset has been judiciously partitioned into distinct training and testing subsets. Vigilance against overfitting has been consistently maintained throughout the training phases of both models. Validation splits have been strategically incorporated to assess the model's performance during training. These considerations lead us to posit that an augmentation and diversification of the dataset will likely contribute to a noteworthy improvement in model performance in terms of Recall and Precision and F1 score.

In the future, stress detection could become more common in daily life. It might be used in things like smartwatches or phones to help people know when they're feeling stressed. In the future, stress detection technology has broad applications that span various sectors. It holds potential in healthcare for remote patient monitoring, enabling real-time insights into mental well-being, and supporting early intervention by healthcare professionals. In workplaces, stress detection can be integral to employee well-being programs, fostering healthier environments and optimizing productivity through workload adjustments. Educational settings may benefit from stress-aware technology, aiding student support services and enhancing adaptive learning experiences. Additionally, stress detection's role in automotive safety, public security, consumer technology, and research signifies a future where it becomes an essential tool for promoting well-being and understanding stress factors across diverse populations.

References

1. S. Koldijk, M. Sappelli, S. Verberne, M. Neerincx, & W. Kraaij, "The swell knowledge work dataset for stress and user modeling research", In Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14). Association for Computing Machinery, New York, NY, USA, Nov. 12, 2014, pp. 291–298.
2. S. Koldijk, M. A. Neerincx and W. Kraaij, "Detecting Work Stress in Offices by Combining Unobtrusive Sensors," in IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 227-239, April-June 2018.
3. P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, & K. Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection." In Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18). Association for Computing Machinery, New York, NY, USA, Oct. 2, 2018, pp. 400–408.
4. P. Sarkar and A. Etemad, "Self-Supervised ECG Representation Learning for Emotion Recognition," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1541-1554, 1 July-Sept. 2022.
5. Y. Dev, M. Namdev, R. Shrivastava, R. Srivastava, "LSTM Based Mental Stress Level Detection using Wearable Sensor Devices", Current Trends in Technology & Science, vol. 11, no. 1, pp. 1–4, Jan 2022.
6. V.T. Ninh, et al. "An Improved Subject-Independent Stress Detection Model Applied to Consumer-grade Wearable Devices." International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Cham: Springer International Publishing, July 2022, pp. 907-919.