Physical Fatigue Detection with Wearable sensors using Machine Learning and Deep Learning Models

Submitted as Research Report in SIT723

Febuary 11, 2022

T3-2021

Gayathri Ramya Sri Chevuru

STUDENT ID 219096729

COURSE - Master of Data Science (S777)

Supervised by: Dr. Sutharshan Rajasegarar and Prof, Dr. Maia Angelova Turkedjieva

Abstract

In professional construction, sports, transportation and industries wearable sensors are increasingly being applied to regulate fatigue. Physical fatigue is a challenging and important workstation ”problem” in production since it reduces productivity and raises the risk of accidents. As a result, physical fatigue must be addressed. This research has two key objectives. First, we look at how wearable sensors can be used to identify physical weariness in industrial simulated operations. Secondly, the purpose is to calculate the level of physical fatigue in the participants with time. An experiment was performed with the help of wearable sensors by considering 8 participants from the workplace. The fatigue levels within the participants are measured by the ratings of perceived exertion (RPE) values. In order to obtain accurate prediction for the physical fatigue detection in the workplace, a machine learning model and a variety of deep learning models were compared (SVM, LSTM, Bi-LSTM, GRU, and Bi-GRU). The results suggest that the SVM and BiLSTM models were efficient at detecting physical fatigue.

Keywords: Wearable sensors, Physical Fatigue, Machine Learning, Deep Learning, SVM, LSTM, GRU, BiGRU and BiLSTM.

Contents

List of Figures

1 Introduction

Workplace fatigue is a complex and multidimensional issue that affects a worker's productivity. It is caused by excessive physical activity and is linked to psychological, economical, and environmental factors. Fatigue has many short- and long-term consequence. Fatigue in the workplace has various factors to be considered along with the mental and physical health. Physical fatigue is defined as a decrease in ability to complete a physical task as a result of previous physical activity. This is a very serious factor especially in the manufacturing industries. This within the workers can affect the performance, quality, quantity, injuries can happen and accuracy decrease. Fatigue can also affect the time taken to complete a task, work rate, and the intervals taken where all these can reduce the manufacturing industry capability. At the same time the actual reason for fatigue is unknown. Research was made in order to identify the exact parameters to prevent from these happening.

Physical fatigue detection currently relies on fitness tests to determine whether a worker has sufficient capacity before start of work, sleep patterns, or brain functioning or muscle performance. The growing availability of sensing technologies, such as wearable sensors can be used to detect the fatigue in the workplace. These sensors can be attached to worker and the pulse readings, sensor reading etc., can be noted and can identify whether the participant is fatigue or not. Wearable sensors are currently less expensive, simple to use, and easy for the worker to wear. They've evolved into the most used devices for tracking mobility and physical activity. There have been few occupational applications directly related to physical fatigue detection in most physically demanding occupations, such as construction, and manufacturing. There is no exact measure for the physical fatigue because of which the heart rate readings were considered and detected based on these readings so far. Furthermore, all these measures are very important as the physical fatigue in the workplace is dangerous and prevention measures can be taken beforehand where the factories can take care before any risk or hazards happening [1].

The main aim and objective of this project is to develop a best physical fatigue detecting model for industrial purpose. The study has been made and few models were created for detecting physical fatigue with the use of affordable wearable sensors. The following three research queries were addressed in order to reach this goal:

(1) What are the best metrics for measuring physical fatigue detection?

(2) Which sensors are good for detecting the physical fatigue?

(3) Which model is the best fit for the physical fatigue detection in the workplace?

2 Literature Review

Physical fatigue within the workers in the workplace is measured using various methods. In this paper it is discussed how the physical fatigue is detected using wearable sensors with the help of various literature research. In this literature study, the physical fatigue was measured by using wearable sensors and the sensor data was recorded for eight individuals who were healthy. Penalized logistic and multiple linear regression models were used in order to detect physical fatigue and estimation level, respectively. The physically fatiguing tasks were divided into Parts Assembly Task, Supply Pickup and Insertion task and Manual Material Handling. Penalized logistic regression for physical fatigue detection and development was performed in correspondence to Random Under Sampling (RUS) and Random sampling. Therefore, the Least Absolute Shrinkage and Selection Operator (LASSO) model with RUS sampling is the best option for modeling physical fatigue detection. But the limitation was the sample size is very small. The effect of demographic variables needs to be explored in future models of physical fatigue [1].

In another similar research, the purpose was to develop a methodology to objectively classify subjects' fatigue level in the workplace utilizing the motion sensors embedded in the smartphones. They have considered 24 participants with a smartphone that was attached to their shank and then asked to squat and gait. Based on Ratings of perceived exertion (RPE) machine learning model Support Vector Machine (SVM) was developed to classify individuals' gait into two (no-fatigue vs. strong-fatigue) and four levels (no-fatigue, low-fatigue, medium-fatigue, and strong-fatigue). The models provided the accuracies of 91 percent and 61 percent respectively. This model is used to detect fatigue and no-fatigue among the workers. The limitation of this model would be what if the fatigue worker is classified as a no-fatigue worker. This could be risk for the workers. So further studies must be done to overcome this, and more classifications needed to be considered for accurate values [2].

Another study uses heart rate measurements to provide an advanced analytical approach for detecting physical fatigue among workers. This paper implemented various machine learning algorithms like Random Forest, Linear Regression, Knearest neighbors (KNN). Amongst all, the KNN model has best accuracy. But this experiment is confined to only small sized datasets. More data could be collected in the future to account for individual variance and to make the more accurate forecasts more generalizable [3].

In addition to the above machine learning models, in the further literature research for fatigue detection, the most common algorithms used in this field were Bayesian learning, SVM, Clustering, Regression, Classification and another subfield of AI is Deep Learning (DL) that uses Artificial Neural Networks (ANN). Here the SVM resulted in 98 percent of accuracy whereas ANN resulted in 99.5 percent of accuracy. The method's limitations could include the need for more sensors to capture EEG data, as well as the inability to connect them to the human body due to noise and movement. The future work was based on two aspects: the detecting technology and its implementation. The approaches for detecting fatigue that are now in use rely on a single type of parameter. Fatigue, on the other hand, is a complicated phenomenon that cannot be effectively represented or recognized with a single feature [4].

Besides fitting machine learning algorithms, deep learning models were also fitted and the research on this indicates more detailed results. In this research, the physical fatigue was measured using deep learning model named Long Short-Term Memory (LSTM) algorithm. In here the accuracy was very low as the feature extraction was very limited due to a very small dataset. The accuracy obtained for fatigue detection was 43 percent. Since limited training was given, it couldn't provide with good results. But this model could be worked on a very large datasets where the training is sufficient, and more features could be extracted for more better and accurate results [5].

In here, the future work of this paper was experimented by feeding a very large datasets to LSTM model and continued to work further to compare the results. This literature review intended to show that the lower body motions and heart rate both contribute to critical information for detecting fatigue. This study was carried out with 60 healthy participants where three levels that are low, moderate and high were estimated from 32 sit-to-stand extracted kinematic features. Random Forest Model was used and an accuracy of 82.5 percent is acquired through it. Because all the participants were healthy persons, one limitation of this study is that the features may generate different patterns and behaviors in patients or other people with varied physical situations. These traits should be used in future research to construct robust models [6].

In the literature research its observed that SVM model provided more accurate results among the machine learning methods and a deep learning model LSTM was implemented for a small dataset. In this paper, some of the future works of the previously reviewed papers were implemented by designing a LSTM model for a larger dataset and experimented with other deep learning models like Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), Bidirectional GRU (BiGRU) with the help of the wearable sensors and the heart rate values that are the Ratings of Perceived Exertions (RPE) for their respective sensors. The below attached is the comparison table constructed for the literature research made [Figure 1]. This entire literature review summary table is available in the Supplementary Resources Section.

JOURNAL	AUTHORS	TITLE	YEAR	LINK	FEATURES	ML / DL ALGORITHM	ACCURACY ACHIEVED	LIMITATIONS	COMMENTS	DATASET USED LINK
IEEE XPLORE	Zahra Sedighi Maman a, Mohammad Ali Alamdar Yazdi a, Lora A. Cavuoto b, Fadel M. Megahed	**A data-driven approach to modeling physical fatigue in the workplace using wearable sensors**	2017	https://deakin365.sharepoint.com/:b:/r/sites/T3-2021-SIT723-ProjectsVeeraVishnandGayathri-Maia/Shared%20Documents/General/Gayathri%20Chevuru/my%20research%20papers/A%20data-driven%20approach%20to%20modeling%20physical%20fatigue%20in%20the%20workplace.pdf?csf=1&web=1&e=NMWAsF	Acceleration related features and jerk-related features	Penalized logistic regression	80.30%	The sample size is very small. The effect of demographic variables needs to be explored in future models of physical fatigue.	Random Under Sampling standard regression LASSO model is implemented	https://githu b.com/zahra me/Fatigue-modeling.git
BioRxiv	Swapnali Karvekar, Masoud	**A Data-Driven Model to Identify Fatigue Level Based on the Motion**	2019	[illegible]	2 level-no fatigue and strong fatigue; 4 level-	SVM algorithm is used for 2 level	91% for 2 level model and 61% for	if a fatigued person is identified as a no-fatigued one, i.e., 47 cases in	FATIGUE DETECTION:An experiment was conducted to collect the motion data from a smartphone during the	https://www xsens.com/n ews/xsens-shares-datasets-

Figure 1: Literature Review Comparison Table

3 Methodology

The key methods used in this research were SVM machine learning algorithm and deep learning algorithms LSTM, GRU, BiLSTM, and BiGRU. This approach combines four phases which are Data collection, Data Pre-processing, Model Fitting and Model Testing. Each phase is explained in detail in their respective section in this paper.

3.1 Data

In this paper, as shown in the Figure 2, eight individuals were considered where two among them were already the factory employees and the rest were the students who work casual at this industry. All the approvals were taken for the experiments to be performed. Every participant's health condition was good for the procedures to be continued. Each participant was subjected to perform two different tasks which were

Parts Assembly (ASM) and Manual Material Handling (MMH). During the exercise, every participant was equipped with four different sensors at four locations on the body (Ankle, Wrist, Hip, Torso). The RPE values were noted during each task and for each sensor [1]. In this experiment participant P1 was considered for training and all the other participants were subjected to testing and like this 56 combinations were executed. And the results were studied. Besides this, P1 and P2 were trained and remaining were tested in another combination.

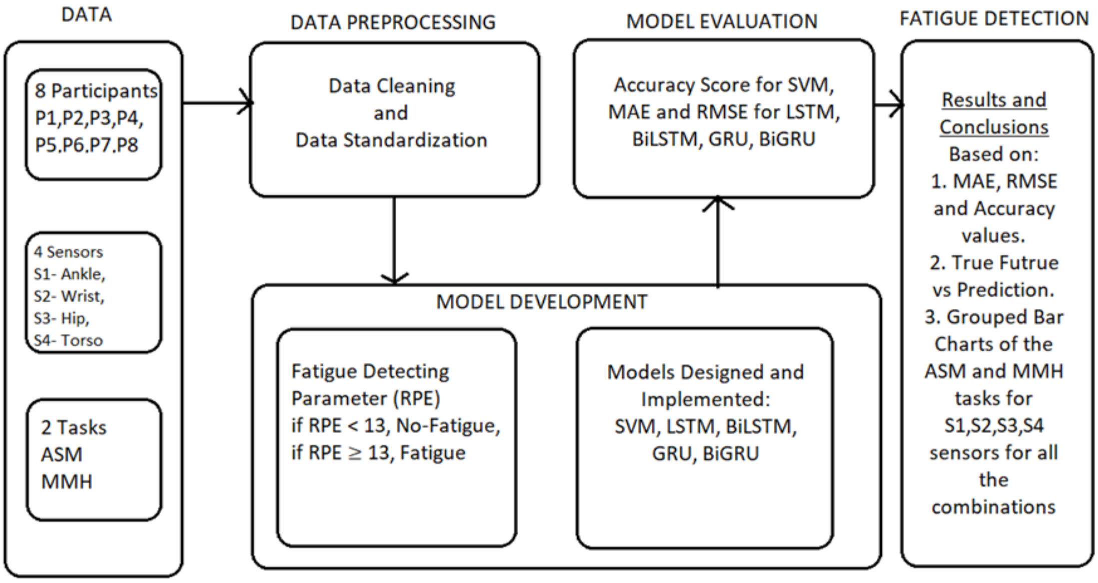

Figure 2: Proposed Model Design.

3.1.1 Equipment

Wearable sensors were utilized here to detect the physical fatigue among the participants. Four inertial measurement units (IMUs) sensors were attached to the participants at four different locations on their bodies. In Figure 3, S1 is sensor at Ankle, S2 is sensor at Wrist, S3 is sensor at Hip and S4 is sensor at Torso were the locations at which the sensors were placed for each participant and the RPE values were recorded in a timely manner for fatigue detection. All the tasks were performed by carrying these sensors throughout the experiments.

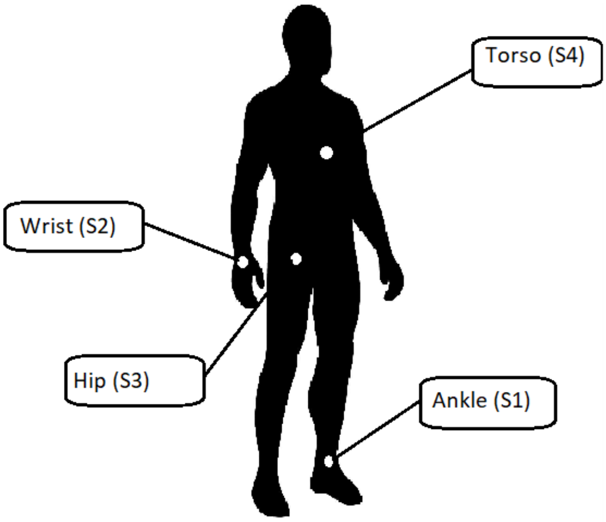

Figure 3: Sensors (S1,S2,S3,S4) and its Locations.

3.1.2 Experiment

The experimental procedure consists of two tasks: Parts Assembly and Manual Material Handling. Each individual was instructed to perform each experiment for a three-hour period with few intervals of 10 minutes in between the tasks. The sensor readings were noted for every ten minutes and obtained the RPE values. The ASM task was performed by assembling the parts by standing at a stationary position. Though this task involves only standing it could be painful as it's a long time stand and a repetitive task. Many industry workers complained about the back pain, hands pain, shoulder pain, fatigue in the legs by standing continuously, and many more. All together can cause physical fatigue within the workers performing such tasks for a very long time and can also affect their performance. Hence this task was considered to detect the fatigue levels. In addition to this, MMH task involves a lot of walking and heavy lifting (weights more than 10kgs and up to 30kgs) which could be the main reasons for the fatigue. And since everything works at a fast pace and this task could be really challenging where the workers continuously involve in walking, standing, heavy lifting and pushing. Due to these tasks the workers experience physical fatigue. Repetitive tasks of such can cause back pain, joint dislocation, legs pain, shoulder pain, and many more. Excessive of these tasks can affect the health and performance of the workers. Hence the physical fatigue is detected by recording the RPE values for every ten minutes for each task for a three-hour duration. Here the physical fatigue levels were detected from the RPE values [1].

3.2 Data Pre-processing

The dataset considered here is a large data consisting of many values for a three-hour period. Initially the data seemed to be unclear and unorganized. Many combinations and trails were made to organize the data into a better format and formed separate spreadsheets for each participant and for each sensor for their respective tasks. Eventhough the data is organized there are many null values within the datasets. So, all the null values were calculated and replaced with zeroes. Finally, the dataset is ready for implementation and model fitting. Prior to the fitting the entire dataset was standardized for the system to understand and analyse the data properly and provide error free best results. In this experiment, 56 combinations were performed by combining different participant's data as training and testing. All of these were subjected to various machine and deep learning models.

3.3 Model development

The main goal of this research is to find out the physical fatigue using wearable sensors within the workplace. In this model, the fatigue is identified with the help of a threshold in the individual's RPE values. Based on the Borg scale, which is the measure of the breathlessness while doing physical work or activity [1], it is stated that:

$$\text{If RPE} < 13\text{, Non-Fatigue, and , If RPE} >= 13\text{, Fatigue.} \tag{1}$$

The above method is applicable for all the sensors and have drawn the results using this Borg scale. All the participants were measured while performing each task and the RPE values were calculated based on the above equations to identify if the worker was physically fatigue or not. With the help of the obtained readings, the datasets were now fitted into various models to predict the future values and compare with the true values. Hence the results were obtained by training and testing the participant's data. In this paper, SVM, LSTM, BiLSTM, GRU and BiGRU models were designed, and the results were compared.

Support Vector Machine (SVM): SVM is a supervised machine learning model used for both classification and regression problems. In this algorithm, each item can be plotted in n-dimensional space (n is number of features) and then the classification is performed with the help of the hyper-plane that perfectly separates the classes into two. Majorly there are two types of SVMs: Simple SVM and Kernel SVM. It has numerous real time applications such as Face Detection, Image Classification, Categorization of text and hypertext and many more [7]. In the below Figure 4, the SVM architecture was displayed.

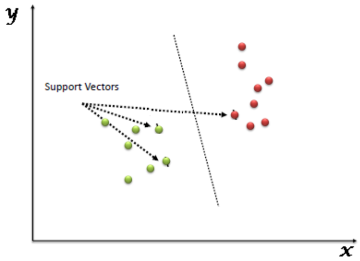

Figure 4: Support Vector Machine Pattern [7]

Long Short-Term Memory (LSTM): LSTM is an artificial recurrent neural network which is implemented in the deep learning field in order to solve the time series-based challenges. This model consists of various gates like input, output and forget which facilitates to remember for a very long time. It has numerous applications like controlling robot, prediction of time series, recognizing speech, forecast of the short-term traffic and many more [8]. In the below Figure 5, the LSTM architecture was displayed.

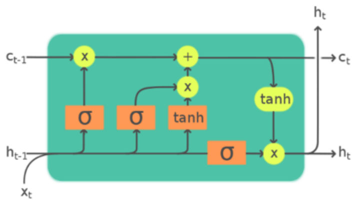

Figure 5: Long Short Term Memory (LSTM) structure [8]

Gated Recurrent Unit (GRU): GRU is derived from LSTM which are similar to each other. One difference is that GRU lacks the output gate. GRU stands out from LSTM as less training parameters and less memory was used by GRU and the execution is faster than the LSTM. But still the accuracy of LSTM remains to be more than that of GRU. In the below Figure 6, the GRU architecture was displayed. River water forecasting could be one of the major applications of GRU [9].

Bidirectional LSTM (BiLSTM): Bidirectional LSTM a model of sequence processing which has 2 LSTM algorithms. As the name suggests, it has an input in forward and other in backward directions. This can improve the information amount that is accessible to the network and can also hold the input explicitly. Radiologically

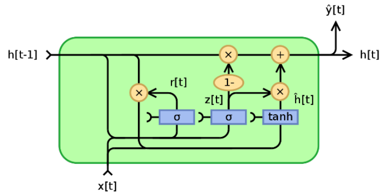

Figure 6: Gated Recurrent Unit (GRU) structure [9]

language modelling using BiLSTM is one of its major applications. In the below Figure 7, the BiLSTM architecture was displayed [10].

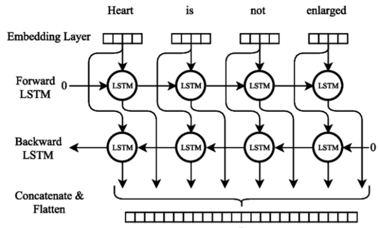

Figure 7: Bidirectional LSTM structure [10]

Bidirectional GRU (BiGRU): BiGRU is also a model of sequence processing which consists of 2 GRUs. In here, the inputs were given in both forward and backward directions. This network has only the input gate and the forget gate. Classification of the emotion from distorted speech is one of the applications. In the below Figure 8, the BiLSTM architecture was displayed [11].

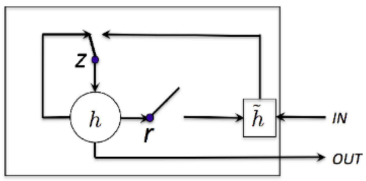

Figure 8: Bidirectional GRU structure [11]

3.4 Model evaluation

Physical fatigue datasets were downloaded from the Github provided by [1]. As shown in the Figure 2, there were 8 participant details and datasets for each task and each sensor for every individual. As mentioned in the above sections that the data downloaded was unorganized which was then prepared to continue further experiments and then the data was cleaned. Later, data was standardized such that the machine can understand for better results to be obtained. Next, the data is set to be implemented and designed for the desired models. Here the dataset was first used to implement SVM model in order to classify whether the person is Fatigue or No-Fatigue. The experiment was the same that the participants were asked to wear the four sensors (Ankle, Wrist, Hip and Torso) while performing both the ASM and MMH tasks, and their RPE values were noted. Here, the datasets were in time series where it was not eligible for splitting the dataset into training and testing. Hence, different participant's data was to be trained and others to be tested. For instance, let's consider P1 was given for training and P2 for testing, and with this combination one SVM model was implemented.

Likewise, there were nearly 56 combinations which were performed to meticulously observe the changes/variations. And finally, the accuracy was evaluated for each combination, and all put together the average accuracy was calculated as an outcome. Besides the machine learning model implementation, a variety of deep learning models were designed in this experiment. Firstly, LSTM model was implemented for a variety of combinations like mentioned above. And here the regression metrics Mean Absolute Error (MAE) and Root Square Mean Error (RMSE) were calculated to identify the model performance. Similarly, other models like GRU, BiLSTM and BiGRU were designed and a variety of combinations of participants datasets were fed into each model separately and the regression metrics MAE and RMSE were calculated. Finally, all the MAE and RMSE values for the 56 combinations of datasets were noted and a comparison table was created. Based on these results the conclusions were processed.

4 Design of the Experiment

In this project, the datasets were downloaded from [1] in order to detect the physical fatigue. The datasets were cleaned and standardized for a better understanding and implementation of various models. The machine learning model SVM and the deep learning models LSTM, BiLSTM, GRU and BiGRU were experimented in this project to detect the fatigue within the participants.

In SVM, accuracy score was measured to validate the model performance and in the other deep learning models (LSTM, BiLSTM, GRU and BiGRU), a comparison table for 56 combinations and for various tasks and sensors was constructed for the values of MAE and RMSE. The design of the experiment is implemented in the same manner as shown in the workflow diagram Figure 9.

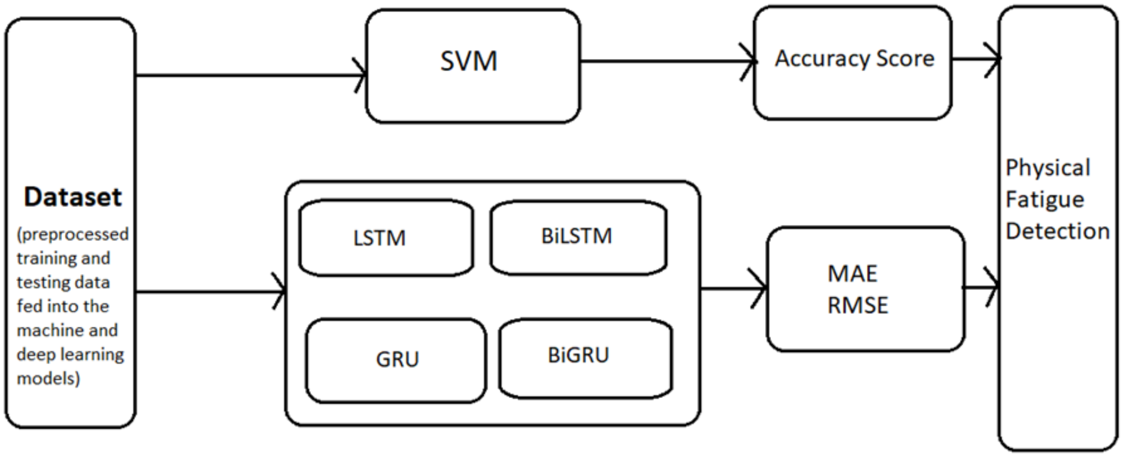

Figure 9: Workflow Design of the models implemented.

4.1 Evaluation Measures

In this research project, a variety of evaluation measures were formulated in order to identify the performance of the models designed for fatigue detection experiment. For the SVM machine learning model, Accuracy score was measured to identify the performance level whereas for the LSTM, BiLSTM, GRU and BiGRU models, MAE and RMSE values were measured and compared. The detailed explanation and their formulas are as follows: Accuracy Score: Accuracy score is one of the classification metrics for identifying the performance of the classification algorithms. To state accuracy in a formulated way, it is defined as [12]:

$$\text{Accuracy score} = \frac{Number\ of\ True\ Predictions}{Total\ Number\ of\ Predictions}$$

Mean Absolute Error (MAE): In order to calculate the Mean Absolute Errors, it is first required to measure the Absolute Error which is the error calculated within the values obtained. Absolute error is basically the difference between the predicted and the true values. Like the name suggests, MAE is the mean of all the absolute errors. The formula for MAE is shown below in Equation (2) [13]:

Root Mean Square Error (RMSE): RMSE is a regression metric used to measure the error obtained in the model while predicting the data. Firstly, the Mean Square Error is calculated by summing the squares of the variance between the true and the predicted values which is then averaged, that can show the deviation in the predicted and the true values. The formula for RMSE is given below in Equation (3) [14]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i| \tag{2}$$

$$RMSE = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y_i - x_i)^2} \tag{3}$$

Where: n = number of errors, y_i = measured values, x_i = true values.

5 Experimental Results and Analysis

In here, it was focused on the results obtained from the experiment made to detect the physical fatigue and these results were analysed based on their classification and regression metrics. In order to obtain more accurate results, a lot of combinations of training and testing datasets were considered. Since there were 8 participants (P1, P2, P3, P4, P5, P6, P7 and P8), 4 sensors (S1, S2, S3 and S4) and 2 tasks (ASM and MMH), many combinations were executed. For instance, when P1 was trained then P5 was tested and likewise when P1 and P2 were trained then P3 was tested. So, in this project a lot of combinations were experimented to study the results clearly. In this section, the conclusions were made based on which task, which sensor, and which model gives the best results. All the codes and the results were added into the Supplementary Resources section.

Initially, a sample data was fed to a machine learning model Support Vector Machine by classifying the data based on the RPE values given in Equation (1). Then the SVM model was implemented, and the classifier metric Accuracy Score was determined for one participant, one task and one sensor. Here 99.95 percent of accuracy for the provided dataset was obtained. This states that detecting whether a person is Fatigue or Non-Fatigue is 99.95 percent accurate using this model. Hence the physical fatigue among the workers within the workplace was detected using this method. Eventhough SVM has given good accuracy, it may have some limitations as here only the RPE values were considered but when other factors and parameters like age, sleep pattern, strength, medical history of the participants, sensor performance, etc., were considered the accuracy might affect. SVM can only classify whether the worker is Fatigue or Non-fatigue but cannot give detailed information about which sensor and which task is giving best results. Therefore, to overcome these issues, a variety of deep learning models were implemented in order to detect the fatigue within the workplace in a participant.

Secondly, entire dataset was used to study and analyse the performance of the four machine learning models (LSTM, BiLSTM, GRU and BiGRU) while detecting the physical fatigue. Here, a lot of combinations were experimented like the participant P1 data considered as training and participant P2 data as testing for one sensor S1 and one task ASM (e.g. P1P2ASMS1). Likewise, considered all the participants data for different sensors (Ankle, Wrist, Hip and Torso) and different tasks (ASM and MMH). In the same manner many combinations were experimented in this project as shown below Figure 12.

ASM				MMH			
S1	S2	S3	S4	S1	S2	S3	S4
P1P2	P1P2	P1P2	P1P2	P1P2	P1P2	P1P2	P1P2
P1P3	P1P3	P1P3	P1P3	P1P3	P1P3	P1P3	P1P3
P1P4	P1P4	P1P4	P1P4	P1P4	P1P4	P1P4	P1P4
P1P5	P1P5	P1P5	P1P5	P1P5	P1P5	P1P5	P1P5
P1P6	P1P6	P1P6	P1P6	P1P6	P1P6	P1P6	P1P6
P1P7	P1P7	P1P7	P1P7	P1P7	P1P7	P1P7	P1P7
P1P8	P1P8	P1P8	P1P8	P1P8	P1P8	P1P8	P1P8

Figure 10: Participant Combinations.

All these 56 combinations were fed into LSTM, BiLSTM, GRU and BiGRU in order to analyse the results and find out which model is the best fit to detect the physical fatigue among the workers in the workplace. The regression metrics MAE and RMSE were calculated. A comparison table of MAE and RMSE values was constructed for all the four deep learning models, four sensors, two tasks and all participant combinations [Figure 13].

Besides this, the time series plots with the time on x-axis and RPE values on y-axis for the True Future and Prediction values were plotted for all the combinations [Figure 14]. From this comparison table and the plots, it can be concluded that the assembling task ASM provided more accurate results than the MMH task for all the models where the predictions were near to the true future. Since the ASM task is a bit easy and there were not many movements, the attached sensors were sufficient to capture the movements and reflect on the readings. But the MMH task involves a lot of walking, bending, lifting heavy, pushing, rotating, etc., these sensors were not sufficient to record the movements effectively. Hence, the MMH task predictions were not accurate. It is also observed that the MAE and RMSE values were very high for the MMH task when compared to that of the ASM task. This indicates that the MAE and RMSE

			LSTM		BIDIRECTIONAL LSTM		GRU		BIDIRECTIONAL GRU	
			MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
P1P2	ASM	S1	1.5248	1.8857	0.9827	1.0921	1.3183	1.5815	0.5037	0.6049
		S2	1.5066	1.8791	0.6147	0.7078	1.3028	1.5351	0.6147	0.7078
		S3	1.4998	1.8737	0.8032	0.8869	1.2814	1.5220	0.5327	0.6387
		S4	1.1723	1.7175	0.9344	1.0252	1.3369	1.6097	0.5044	0.6057
	MMH	S1	1.7859	2.3726	1.6588	2.2964	1.7361	2.3361	1.7714	2.3365
		S2	1.7656	2.3535	1.6922	2.3113	1.7206	2.3234	1.6768	2.2978
		S3	1.6488	2.2916	1.6680	2.2987	1.7116	2.3238	1.6826	2.2964
		S4	1.7816	2.3694	1.6538	2.2941	1.7199	2.3278	1.6531	2.2946

Figure 11: Comparison Table: MAE and RMSE values for four models for four sensors and two tasks.

values for ASM were more accurate.

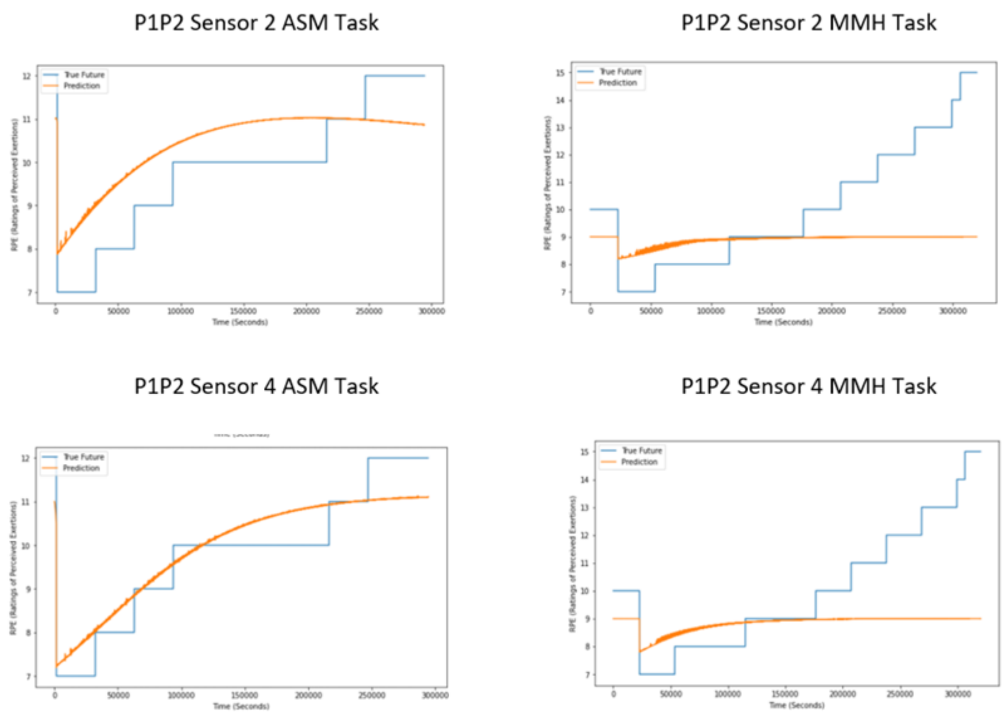

Figure 12: True Values vs Predictions of P1P2 S2 and S4, ASM and MMH tasks.

In addition to this, grouped bar charts were plotted for all the sensors and tasks for each model and combination to compare their performance for all the combinations [Figure 15]. The MAE and RMSE values were noted from all the combinations and all the models executed. Using these values grouped bar charts were plotted for the ASM and MMH tasks and 4 sensors. These Grouped Bar Charts show that the torso (S2) and the wrist (S4) sensors were essential to detect the fatigue. Even from these charts it is easy to extract the information that the ASM is more accurate compared to that of MMH for physical fatigue detection.

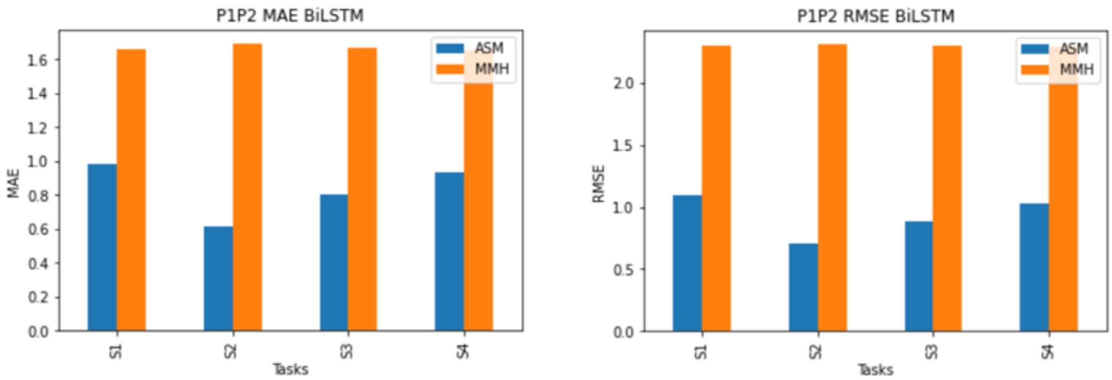

Figure 13: Grouped Bar Charts for P1 and P2 combination of ASM and MMH tasks and S1,S2,S3,S4 sensors.

And it is also observed that among all the other models, BiLSTM gave the best results whereas and BiGRU gave average results but, LSTM and GRU results were not good as the MAE and RMSE values for BiLSTM were more inclined towards zero whereas these values for the other models were far away from zero. As the error is more the performance is less, hence in this way the best method for physical fatigue detection is identified. From the comparison table and time series plots, it is easily understood that the BiLSTM has lower MAE and RMSE values and obtained more accurate results in the True Future vs Prediction plots.

The above results illustrates the BiLSTM outcomes for the participant P1 (train) and P2 (test) for the sensors S2 and S4 and tasks ASM and MMH. From these graphs, it is easy to conclude that for which task is the performance of the model is good and which sensor gives better results that helps to detect the physical fatigue. From this figure itself it is observed that the ASM task provided more accurate results whereas the MMH task doesn't. The less variance of trends between the true future and the predicted RPE levels led to this conclusion.

6 Discussions and Conclusions

In the manufacturing industries it is important to monitor the physical fatigue within the workers as a safety measure. The fatigue detection techniques need to be implemented in every workplace in order to avoid any accidents or injuries from happening for the workers. In this project, a variety of physical fatigue detecting models were implemented. In this project, actual industrial tasks were considered like Part Assembly (ASM) and Manual Material Handling (MMH) in order to detect the fatigue while performing these tasks by the participants. To extract the data of the physical condition of the workers, four sensors were attached to each participant as shown in the Figure 2.

As there were 8 participant's data, 56 combinations of trails were executed like mentioned in Section 5. The SVM model gave good accuracy of 99.96 percent in classifying Fatigue or Not-Fatigue, but this model gives limited information about the sensor's performance for each task and comparisons. Hence, all these combinations were fed into the 4 deep learning models.

From the results obtained from the grouped bar charts, it is observed that the Wrist (S2) and Torso (S4) sensors were more sensitive to detect the fatigue among the workers. Besides this, from the time series plots of the True Future vs Predictions, it can be concluded that the BiLSTM algorithm can be used for modelling in the industries as they gave the best results amongst all the other models. Their predictions were close to the actual values from which it can be proved that the BiLSTM model works well to detect the physical fatigue. But there is a limitation as this model worked well only for assembling ASM task and not for the MMH. Since the ASM task is a bit easy and there were not many movements, the attached sensors were sufficient to capture the movements and reflect on the readings. But the MMH task involves a lot of walking, bending, lifting heavy, pushing, rotating, etc., these sensors were not sufficient to record the movements effectively. From the comparison table [Figure 9], it is observed that the MAE and RMSE values for ASM task and S2, S4 sensors and BiLSTM models were reasonable and lesser compared to other values. For instance, the MAE and RMSE values obtained for BiLSTM for ASM task for sensor S2 in P1P2 combination were 0.6147 and 0.7078 respectively, whereas for LSTM model the MAE and RMSE values for the same combination were 1.5066 and 1.8791 respectively which were with larger error. Hence the BiLSTM model performance is best for all the combinations when compared to other models. Some improvements to be made in considering the sensors and its readings to achieve good results for MMH task as well.

All in all, in this project the physical fatigue detection models were designed and implemented in order to achieve the purpose of this research. It was observed that the BiLSTM model was giving better performance compared to others, but it is not perfectly applicable for the industry purpose at this stage as there are some limitations. From the results obtained, this model wasn't providing good results for MMH task. At the end, a small trail was made after analysing all these combinations that was another type of combinations was experimented by considering two participants for training and one for testing like P1 and P2 were trained and P3 was tested. For this trail, the results were not good at all as the predicted values were nowhere close to the true future values.

So, studies should be made further in order to improve and achieve better results while performing any industrial task by the workers in any combinations of training and testing data of the workers fed into the models. Finally, these were the discussions and conclusions made from this research project. The Python codes implemented and executed in this study were provided in the GitHub and the link is given in the Supplementary Resources.

7 Limitations and Future Work

In this project, there were few gaps that need more attention in order implement these methods in the industries. Firstly, the physical fatigue detection was not good for MMH task using this BiLSTM model. Further studies must be done to improve these results. Secondly, additional sensors needed to be considered to improve the results and to detect the fatigue based on various factors like health conditions, sex, sleep patterns, age, etc. As in this project data, RPE is the only fatigue detecting factor and in real life scenario it's not ideal to rely on just one deciding factor and needed to add other deciding factors too for better performance and further research needed to be done on implementing better detecting techniques. Future work will be discovering and experimenting various sensors for better detection techniques. Thirdly, when the two participant's data was fed into the models as training and another participant's data for testing, the results obtained were not good as the predicted values varies more when compared to the true values. And their MAE and RMSE values were also very high. Some further research study must be made to fill these gaps. Finally, it is recommended to experiment many other combinations, sensors, and tasks and improve the techniques by resolving all the limitations mentioned above.

8 Supplementary Resources

In this project, the Python codes for all the algorithms and obtained results were linked to Github Repository https://github.com/Gayathri-Chevuru/Fatigue-Detection which is publicly accessible.

9 References

1. (Sedighi Maman, Alamdar Yazdi et al. 2017), A data-driven approach to modeling physical fatigue in the workplace using wearable sensors.

2. (Karvekar, Abdollahi et al. 2019), A Data-Driven Model to Identify Fatigue Level Based on the Motion Data from a Smartphone.

3. (Nasirzadeh, F., et al. 2020). Physical Fatigue Detection Using Entropy Analysis of Heart Rate Signals.

4. (Balaskas, K. and K. Siozios, 2021), Fatigue Detection Using Deep Long Short-Term Memory Autoencoders.

5. (Aguirre, A., et al. 2021), Machine Learning Approach for Fatigue Estimation in Sit-to-Stand Exercise.

6. (Hooda, R., et al. 2021), A comprehensive review of approaches to detect fatigue using machine learning techniques.

7. Ray, S. (September 13, 2017). ”Understanding Support Vector Machine(SVM) algorithm from examples (along with code).” Retrieved Febuary 09, 2022, 2022, from https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/.

8. Wikipedia (2022). ”Long short-term memory.” Retrieved Febuary 09, 2022, from https://en.wikipedia.org/wiki/Long-short-term-memory.

9. Wikipedia (2022). ”Gated recurrent unit.” Retrieved Febuary 09, 2022, from https://en.wikipedia.org/wiki/Gated-recurrent-unit.

10. Papers With Code (2022). ”Papers with code - BILSTM explained,” BiLSTM Explained.” Retrieved Febuary 09, 2022, from https://paperswithcode.com/method/bilstm.

11. Papers With Code (2022). ”Papers with code - BIGRU explained,” BiGRU Explained.” Retrieved Febuary 09, 2022, from https://paperswithcode.com/method/bigru.

12. Developers, m. l. c. c. g. (2022). ”Classification: Accuracy.” Retrieved Febuary 09, 2022, from https://developers.google.com/machine-learning/crash-course/classification/accuracy.

13. Stephanie (2020). ”Absolute error and mean absolute error (MAE).” Retrieved Febuary 09, 2022, from https://www.statisticshowto.com/absolute-error/.

14. Moody, J. (2019). ”What does RMSE really mean?”. Retrieved Febuary 09, 2022, from https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e.