Lifelogging - Temporal Human Activity Classification

Aine Mary Harte
Msc in Computing, Data Analytics
ainemharte@mail.dcu.ie
15380396

Ananya Borah
Msc in Computing, Data Analytics
ananya.borah2@mail.dcu.ie
19210250

Joel Peter Henry Rozario
Msc in Computing, Data Analytics
joel.henryrozario2@mail.dcu.ie
19210102

Abin Kuriokose
Msc in Computing, Data Analytics
abin.kuriakose2@mail.dcu.ie
19210152

Ajay Mukundan
Msc in Computing, Data Analytics
ajaymukundan2@mail.dcu.ie
19210980

***Abstract*— Quantified Self – A self-tracking project is a new phenomenon where people use empirical methods to reflect on their own personal decisions. Often referred to as 'Everyday Science', this contemplative ability can improve self-awareness, learn more about oneself and even motivate to change one's lifestyle. This study attempts to find patterns in a lifelogger's daily activity and thus can be used to perform behavioral analysis, as studies suggest. This research also focuses on analyzing the efficiency of classifiers to fragment activities into daily life patterns. Data was gathered by a lifelogger for NTCIR-14 task via two wearable sensors, and a lifelogging camera for 28 days. The study includes clustering/identification of large variability of actions, typically categorized into 15 activities, based on biometrics and location data. Classification of Human Activities was performed with Supervised Learning algorithms, Ensemble learning- Random Forest and Boosting, and on a trained Artificial Neural Network – Multi-layer Perceptron model. Ensemble learning with decision trees and Multi-layer perceptron of Neural Networks classified with a 70% accuracy**

Keywords— Quantified Self, Lifelogging, NTCIR-14, lifestyle pattern, Classification, Supervised Learning, Ensemble Learning, Neural Networks

I. INTRODUCTION

The availability and accuracy of wearable sensor technology have increased substantially in recent years, as have lifelogging devices such as high quality portable wearable cameras. This has sparked an inevitable increase in the number of people participating in self-monitoring using sensor technology [2][3]. This is what Quantified Self is all about. It is quite crucial to self-articulate the reason to self-monitor as it can push oneself to make lifestyle changes, improve dietary habits, motivate to learn new skills, and countless other possibilities. There are countless possibilities where these insights can be applied to e.g. health care, Travel, Psychological Analysis, Finance, and many other domains [5]. Self-knowledge through numbers includes Questioning, Observing, Reasoning, and consolidating Insight as explained by Quantified Self. With the advancement in 'Everyday Science', data stored in self-tracking devices can be turned into meaningful insights. These insights can often bring about a lifestyle change if observed and consolidated properly [10].

These exciting potentials have encouraged us to indulge in analyzing human behavioral patterns from lifelogging data. The main objective of this study is to test the efficiency of analyzing the temporal segmentation of human activities, with an underlying goal to test the predictability of a subject's individual activities. Dataset, including images and features, was captured as a task for NTCIR-14[2]. The wearable sensor captured features such as location, calories burned, steps count, distance traveled and heart rate, glucose level every 15 minutes. Meanwhile, a lifelogging camera captured images every 10 seconds and was trained on Neural Networks to perform feature annotation. The dataset tried to capture activities performed by a lifelogger but managed to classify only 2 activities i.e. Walking and Transport. Our Study focuses on classifying activities throughout a day based on metrics and perform a supervised learning task to classify future activities based on metrics and location.

The supervised learning task is performed through Ensemble learning- random forests and boosting, and Artificial Neural Network – Multi-Layer Perceptron. Source code link is added at the end of this paper.

This research paper focuses on addressing the following questions:

1. Can a person's daily behaviour be determined with lifelogging data and biometrics?
2. With the aid of mobile sensor data and glucose measuring device, apart from visual concepts, can we improvise human activity recognition?

II. RELATED WORK

Ekaterina et.al [1] performed a temporal segmentation and activity classification of cooking a recipe from a first-person camera, which is like a lifelogging wearable camera. The goal was to perform an activity analysis to determine the timeline of activities involved in cooking a recipe. Data was obtained from sensors and a camera, whereas high-level activities were manually annotated. The classification was tested with supervised Hidden Markov and K-NN models. K-NN outperformed HMM due to the high dimensionality of the dataset, even after performing Principal Component Analysis. Even though the results were promising with 94% accuracy, the tests were performed on a small scale and require robust performance on a larger sample which was acknowledged by the researchers.

According to the work done in supervised learning on lifelogging data [2], 27 visual concepts were extracted with probabilities and concept scores from lifelogging images. Support Vector Machine with iterative approach yielded good results on concept detection. The visual feature extraction was based on work from another research. The validation of these annotations was performed manually by the lifelogger's on 95, 907 images and, achieved 75% accuracy in its classification. The paper further explored the temporal segmentation with transitional probability and Pointwise Mutual Information (PMI). The robustness of classifying lifelogged activities can be now improved by extracting much more features with advancements in HOG, Inception V3 features.

A quite interesting study was performed by Daragh Byrne et. al [4] on observing event similarity by collecting GPS coordinates and MAC address of others, deemed friendly, from Bluetooth. The study was carried out by segmenting the images based on the subject's activity on a day-by-day basis, whereas their respective GPS and Bluetooth information were observed to study pattern. It is revealed that high motion activities such as travelling, walking, running was detected at a higher accuracy with CPU, whereas low-level motions were observed quite efficiently with friendly MAC addresses from Bluetooth data. The similarity study can be improved with other metrics such as distance travelled and exact location instead of coordinates.

Work done by Alan Smeaton et. al [3] proposes the use of lifelogging therapy, known as Sensecam Therapy, for stimulating the cognition of a person suffering dementia. Personal Identity was constructed by using the visual concepts in features extracted by Image recognition. A case study on three persons revealed that Sensecam therapy managed to construct a more 'holistic versions of identity' and warranted appropriated protocols for official medical therapies. This study showed the potential of lifelogging in healthcare industries with the introduction of proper protocols deemed necessary for effectiveness

A much more practical approach on the effectiveness of lifelogging was studied by Lindley et. al [5]. With improvements in storage, computing resources, and sensors, progress has been made on eliciting patterns e.g. behavioural and lifestyles, from lifelogged data. This paper explored the possibility of improving lifestyle choices with lifelogged data which was later developed and termed as 'Quantified Self' by a team of enthusiastic lifeloggers. A family was given lifelogging devices for a week's duration and they were shown the result which showed glimpses of positivity. According to the study, though the degree of impact is not determined, researchers believed that the outcome made an influence such that the family made a few positive lifestyle changes.

ZhiWo [6] was developed as a personal archive where activities were classified automatically based on the sensor and other metrics from mobile devices. Support Vector Machine with Linear Kernels was used to segment activities and activity retrieval was based on the timeline. Though this is a prototype, there is no test evidence on the efficiency of this prototype in classifying activities. The usage of SVM for activity segmentation achieved better results as studied previously by Cathal et. al [2]. The developers also added the option of correcting misclassified activities manually.

A research on human activity analysis to identify 22 unique lifestyle traits was done by Cathal Gurrin et. al [7]. A Lifestyle interpreter tool (LIT) was developed to effectively investigate one's lifestyle. Studies like this paved way for the 'Quantified Self' initiative. This work proposed various subjects on the potential applications of lifelogging in the field of healthcare, personal logging, lifestyle improvisation, and psychology. A lifestyle trait Interpreter was modelled and managed to achieve 70% efficiency in determining characteristics trait.

A study by Yi Chen et. al [8] reveals Lifelogging can improve memory augmentation and suggests guidelines for introducing this into the field of healthcare and psychology. A theoretical study on how to improve memorability on certain incidents was done on implicit and episodic memory. Memory cues and information retrieval was the basis for invoking those memories. The study was not concluded with statistically significant results and was done on a smaller scale.

Recognizing human activities based on sensors and visual concepts are still a big challenge in the current landscape. Works mentioned above have contributed immensely to this field of study.

III. DATA MINING METHODOLGY

This research paper has implemented CRoss-Industry Standard Process for Data Mining (CRISP-DM) process. Below is an explanation on how the process model is implemented to achieve the best results.

A. REQUIREMENT ANALYSIS

The main objective was to study the efficiency of analysing temporal segmentation of human activities to understand the subject's daily life pattern from lifelogged data. The focus was to extract information from certain features such as steps taken, distance, location, heartbeat and to classify conventional activities of a life logger as well as to predict his pattern for the next 5 to 7 days. Out of the 24 essential activities for human survival, as published by Daniel Kahneman et. al [10], we have classified 15 basic activities as follows: Sleeping, Working, Eating, Relaxing, and Travelling to a specific location based on his timeline e.g.

Travelling to work, travelling to Relax, Travelling to Home and so on.

As the initial temporal segmentation of activities is done on a mundane level, with further study in this domain, an extensive analysis on Human Activity Recognition can be performed, which can aid in providing recommendations and predictions to the life logger as well as various industries.

B. EXPLORATORY DATA ANALYSIS

1. Dataset – A Brief Introduction

The NTCIR-14 lifelog data consists of data from wearable biometric trackers every minute along with the continual blood glucose data observed every 15 minutes from 21 days. The Dataset consists of 16 features with few substantial attributes like calories burnt, minute id, heart rate, glucose, scanned glucose, activity, location, steps, and distance.

2. Exploratory Data Analysis

Null values were observed in Activity and Heart rate which were incredibly significant for feeding as training data. To impute the missing heart rate and to classify activities based on other independent variables, Pearson's Correlation was calculated. It is found that heart rate has a medium to a strong positive correlation with Steps, Calories, and Distance as shown below.

Fig. 3.2.1 Linear Regression Imputation - A correlation matric of bio metrics

As activities were classified only as walking and commuting, further analysis of metrics was required to classify activities into the 7 categories as mentioned in the previous section. There was a clear seasonality in the lifelogger's glucose pattern from which we could infer the lifelogger's food consumption pattern. A well visible spike can be observed in the subject's continuous glucose level after he/she consumes their respective meal. Steps and Distance are used to find the subject's movement pattern. All the biometric values were normally distributed with skewness to the left which indicates the lifelogger had a sedentary lifestyle.

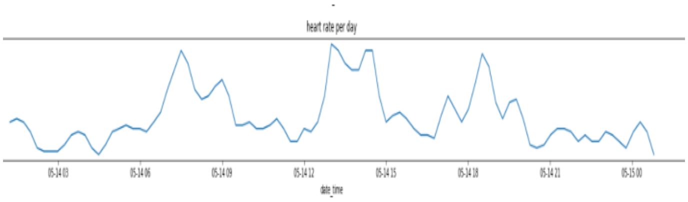

Fig. 3.2.2 Time series plot of the lifelogger's Glucose activity exhibiting trend-cycle over a day.

Latitude, Longitude, and Song attributes were deemed unnecessary as they had too many null values which could not be imputed.

C. FEATURE ENGINEERING

1. Imputation – Multiple Linear Regression

As mentioned in previous section, Heart rate has a medium positive linear relationship with steps, distance, and a strong positive relationship with calories. As data was missing on a random scale, Heart rate was imputed with a simple random imputer followed by multiple linear regression imputation. To avoid multicollinearity, the distance metric was removed as it a had near perfect correlation with steps. The predicted values on the hyperplane introduced little noise to the heart rate data. This could have been improved with Stochastic Regression, but the bias introduced was negligible.

$$\textbf{Heart rate}_{(i)} = 68.3871 + -0.2151 * X_1 + 8.2429 * X_2$$

p-value < 0.05 and the R Squared value is 0.7 approx., which indicates a good fit.

2. Z- Score Normalisation

It can be seen that the observed bio-metrics are of varying scale i.e. heart rate per minute, distance travelled in km, steps walked per minute, etc. To get better classifiers these metrics were normalised with Z-score normalisation as they this standardisation is robust to outliers or new values, which is inevitable in lifelogging data.

3. One hot label Encoding of Categorical variable and DateTime Object conversion for Time series Analysis

The location attribute is a substantial feature for better classification. Being a multi-class classification problem, one hot encoding was performed for the categorical variables. As the paper focuses on temporal segmentation, a date-time object or a timestamp object had to be extracted from the given features. Minute ID is a feature that had timestamps along with the image id as a String object. This feature is split, and the date-time is extracted and formatted appropriately for time series analysis and visualisation.

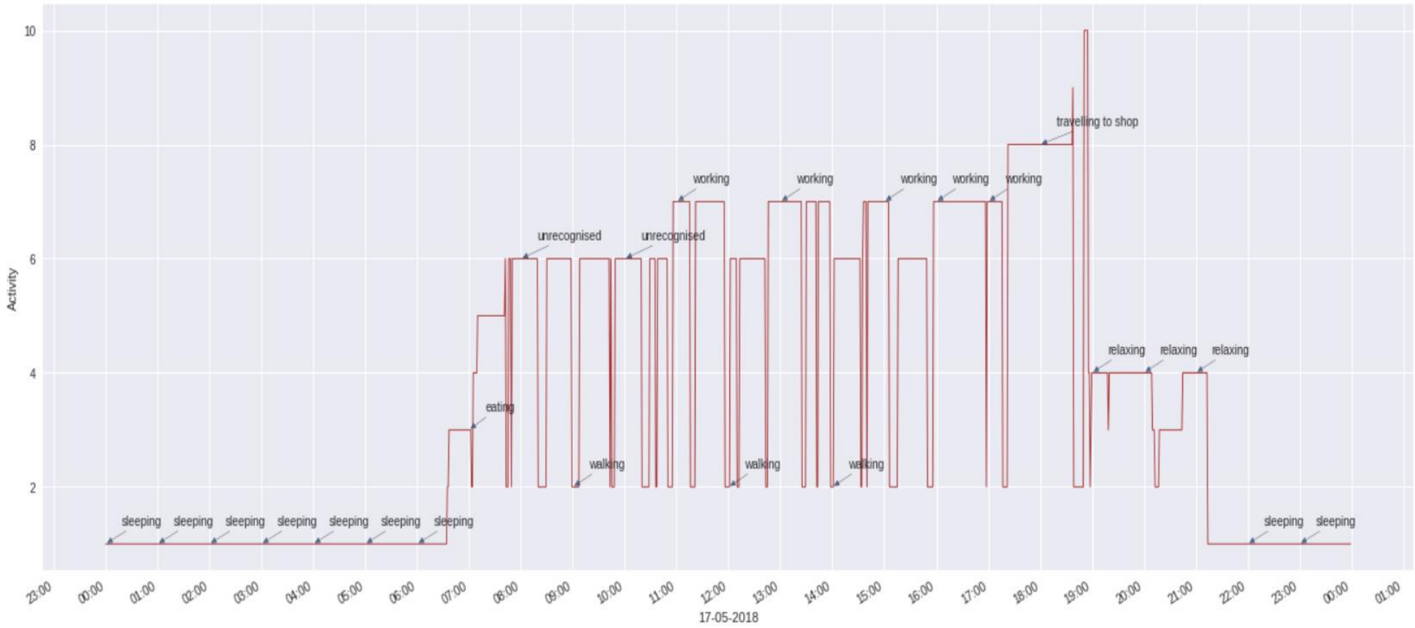

Fig. 3.4.1 A timeline of activities performed by the lifelogger throughout a day.

4. Missing Activity Clustering based on biometrics

In the NTCIR -14 lifelogging dataset, activities were classified only as Transport and Walking. In order to train the classifier, we had to extract features from other metrics such as location, Steps, Distance, Heart rate, and Glucose and label all 15 activities. The NTCIR dataset had lifelogging images and bio metrics. Images were taken from the wearable camera when the lifelogger wakes up and starts his routines. So, the period for when the lifelogging images are unavailable is classified as Sleeping. The Eating activity was classified based on the Glucose spike and cross-validated manually with the lifelogging images.

$$DiffDur = |Duration(X_i) - Duration(Y_i)|$$

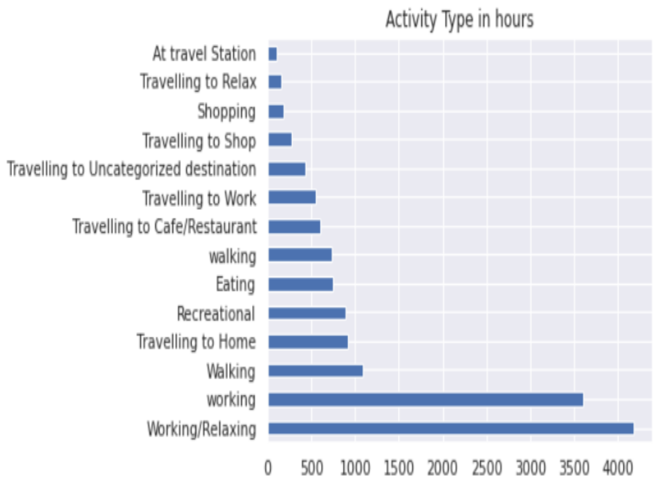

Fig. 3.4.2 Activities classified based on lifelogging data.

Transportation of the lifelogger was classified based on the location, whereas, working activity were classified based on the lifelogger's location and walking pattern.

IV. EVALUATION AND RESULTS

All the models will be evaluated with an F-score accuracy measure as the data is imbalanced. F-score test is a measure of the harmonic mean of Precision and Recall. F,

$$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

A. CLASSIFIER MODELS

1. Ensemble Learning - Random Forest Classifier, Decision Trees

Random Forest has been proven to be an extremely powerful tool, which produces accurate results with regards to classification [11]. The Random Forest method has also been used to make predictions based on time series data with some success [12]. A Random Forest was created for this study using the attributes steps, calories, glucose, distance, and imputed heart rate and location's name (one-hot encoded) as the features. The activity column was then used as the target label for the Random Forest Classifier. The model was trained with 70% lifelogged data, which was approximately 2 and a half weeks.

Model	*f1-score*
Random Forest Base Classifier	0.7317
Grid Search - Best Classifier	0.736976028
10-fold Cross-Validation	0.708130565

Fig. 4.1.1 Precision Accuracy of Random Forest Model

Feature Importance revealed Glucose has low significance on the activity classification precision. Therefore, Glucose was removed, and to find the best hyperparameters for the classifier, a Grid Search was performed which ultimately had a small increase in model performance. The best performing hyper-parameters were estimators - 800, maximum features - 3, maximum depth - 150, minimum sample leaves – 3, and minimum samples split – 10. An increase in the number of estimators from 500 to 800 and maximum feature from 3 to 4 had led to this increased accuracy.

To ensure that the data did not dive into the pitfall of increased variance/Overfitting of the training data, the tuned classifier was 10—fold Cross Validated, and the model classified activities at a mean 70% accuracy measure.

Random Forest Classifier - 10 fold Cross Validated				
Activity	**precision**	**recall**	**f1-score**	**support**
At travel station	0.97	1	0.98	29
Eating	0.5	0.05	0.09	236
Recreational	0.95	0.97	0.96	266
Shopping	0.83	0.79	0.81	56
Sleeping	0.83	0.93	0.88	2622
Travelling to cafe/restaurant	0.13	0.02	0.03	180
Travelling to home	0.22	0.07	0.11	286
Travelling to relax	0	0	0	35
Travelling to shop	0	0	0	82
Travelling to uncategorized destination	0.25	0.06	0.1	119
Travelling to work	0.35	0.04	0.07	177
Unrecognised	0.37	0.87	0.52	492
Walking	0.83	0.66	0.74	539
Working	0.95	0.99	0.97	1100
Relaxing	0.64	0.64	0.64	1248
Accuracy			**0.74**	**7467**
Mean AUC Score of 10 fold CV - Random Forest: 0.7081305654849182				

Fig. 4.1.2 Classification Report of 10-fold Cross Validated Model

Few activities were well classified comparatively. Moreover, activities such as, at travel station, Sleeping, Recreational, Walking, and Working were classified better, and it is can be perceived that it is due to better labelling of Location data.

2. *Ensemble Learning – Boosting, AdaBoost with Logistic regression*

Boosting refers to a set of algorithms for improving model predictions by converting weak learners to strong learners. It trains the model sequentially by learning from the mistakes of previous learners [13]. AdaBoost is one such algorithm with decision trees with a single split called stumps. Logistic Regression can be boosted by using AdaBoost which can select the best model [14]. The model is trained with a Multinomial Logistic Regressor to perform a multi-class classification problem. Both Support Vector Classifier and Logistic Regression, in which logistic regression fared relatively well. A grid search was performed to tune in the hyper-parameters i.e. estimators = 200 from 100.

Model	*f1-score*
Logistic regression with AdaBoost	0.6818
Grid Search - Best Classifier with 5-fold Cross-validation	0.6842

Fig. 4.2.1 Accuracy of Logistic regression with Boosting

The Model was 5-fold cross-validated to avoid overfitting as in the case of the previous method. Adaboost with base estimator as logistic regression performed slightly worse in comparison with random forest decision trees. This can be due to the noises and sparsity of the dataset.

Logistic Regression with AdaBoost with Cross Validation				
Activity	***precision***	***recall***	***f1-score***	***support***
At travel Station	0.71	0.12	0.21	40
Eating	0.23	0.28	0.26	218
Recreational	0.76	0.9	0.82	264
Shopping	0.96	0.45	0.61	60
Sleeping	0.78	0.95	0.86	2597
Travelling to Cafe/Restaurant	0	0	0	165
Travelling to Home	0	0	0	296
Travelling to Relax	0	0	0	41
Travelling to Shop	0	0	0	87
Travelling to Uncategorized destination	0	0	0	133
Travelling to Work	0	0	0	148
Unrecognised	0.36	0.97	0.52	520
Walking	0.83	0.61	0.7	539
Relaxing	0.63	0.31	0.41	1268
Working	0.97	0.95	0.96	1067
Accuracy			***0.68***	***7467***

Fig. 4.2.2 Accuracy of Logistic regression with Boosting

3. *Artificial Neural Networks – Multi-layer Perceptron (3- layer Classifier)*

Multi-layer Perceptron is used as it can distinguish data that are not linearly separable. This architecture has 3 layers: the input layer with 50 nodes, to which all 50 input features are fed to, a hidden layer, and an output layer with 15 nodes to map with the 15 activities that are to be classified. SoftMax activation, a form of logistic regression, outputs a probability of range 0 to 1. Thus, it is much favourable for multi-class classification and used as the firing function for the output layer.

$$\text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \text{ for } j = 1,\dots,K$$

Whereas Rectified Linear unit, ReLu - ***max (0, x)***, is used for the input and hidden layer as it overcomes the vanishing gradient problem in sigmoid function and allows backpropagation which supports the principle of gradient descent. Categorical Cross Entropy is the loss function as it is a multi-class classification problem and the target variables are fed as a one-hot vector. Adaptive Movement Estimation optimizer (Adamax), which automatically tunes the learning rate and decay rate, is the preferred optimizer as the data is categorical and sparse (one-hot encoded). To avoid overfitting of training data, the model was trained with Ridge Regularization, L2, as well as dropout technique. Both the models were outperformed by the model which was trained without L2 regularization and Dropout. This is due to the sparsity of encoded data.

Multi-layer Perceptron ANN Classifier		
Classifier	**CCE Loss**	**F-1 Score**
MLP - Without Regularization	0.694007	74.30%
MLP - L2 Regularization	0.769503	73.60%
MLP Without L2, 5 - fold Cross Validated		70.40%

Fig. 4.3.1 Precision Value of ANN

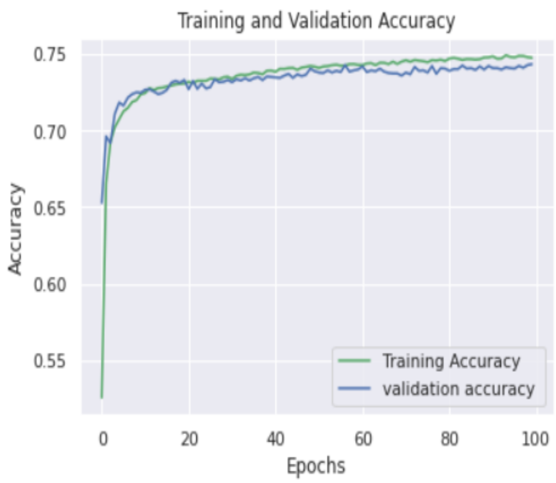

Fig. 4.3.2 Neural networks learns in each iteration. No signs of overfitting/underfitting – Validation and Training Accuracy

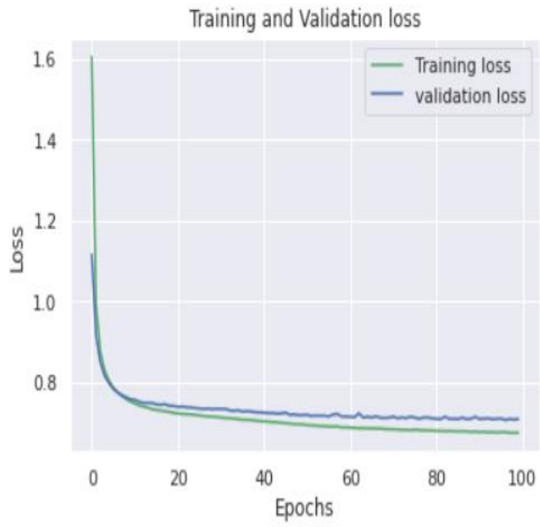

Fig. 4.3.3 Neural networks learns in each iteration. No signs of overfitting/underfitting – Validation and Training Loss

4. *Consolidated Classifier Result*

Model Classifiers - Results	
Classifier	***F-1 Score***
MLP Without L2, 5 - fold Cross Validated	70.40%
Random Forest - 10 fold -Cross Validation	70.80%
Boosting - Logistic Regression 5 fold validation	68.42%

Fig. 4.4.1 F-1 Score of all Cross validated Models

V. FUTURE WORKS

It is quite evident that lifelogging data can be used in many domains such as health care, Technology, psychology, and so on. The research done by us can be improved by better labelling and more research on Multivariate Time series Analysis with Recurrent Neural Networks. Long-short term memory, LSTM, can be used to classify Human Activities and can be used to identify inferable patterns, e.g. Cooking patterns of a lifelogger [1]. Moreover, it can be applied to the prediction of biometrics, for example, glucose level of a lifelogger after his meal, which is not explored yet. With advancements in CNN and feature extraction such as Inception V3 and HOG, Image annotations can be improved, which can contribute significantly to Human Activity Recognition.

Another exciting domain is the identification of activities occurring together [2]. In our research, we have classified an activity as Working/Relaxing when the user is at home watching TV and working on his/her laptop. Research can be done on this with better annotations and further dive into Deep Learning.

VI. CONCLUSION

Cross-validated random forest decision trees, and ANN Multi-Layer Perceptron performed with satisfactory results, whereas Boosting with logistic regression was comparatively low. Decision Trees had slightly better accuracy compared to MLP though. While the results were satisfactory, we feel the performance of these models can be improved with better labelling of activities.

Human Activity Classification with Lifelogging data will present countless scope in the near feature as technology evolves. It is fair to say that we have just merely scratched the surface on the practical applications of this ever-evolving domain.

REFERENCES

[1] E. H. Spriggs, F. De La Torre and M. Hebert, "Temporal segmentation and activity classification from first-person sensing," 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Miami, FL, 2009, pp. 17-24.

[2] Byrne, Daragh, Doherty, Aiden R., Snoek, Cees G. M., Gareth J.F Jones and Alan Smeaton, "Everyday concept detection in visual lifelogs: validation, relationships and trends," 2009 Multimedia Tools and Applications, 49 (1). pp. 119-144.

[3] Piasek, Paulina & Irving, Kate & Smeaton, Alan. (2014). Using Lifelogging to Help Construct the Identity of People with Dementia

[4] Daragh Byrne, Barry Lavelle, Aiden R., Doherty, Gareth J.F. Jones and Alan F. Smeaton, "Using Bluetooth and Gps Metadata To Measure Event Similarity In Sensecam Images," 2007 Information Sciences, pp. 1454-1460

[5] Doherty, Aiden & Hodges, Steve & King, Abby & Smeaton, Alan & Berry, Emma & Moulin, Chris & Lindley, Siân & Kelly, Paul & Foster, Charles. (2013). Wearable cameras in health: The state of the art and future possibilities. American journal of preventive medicine. 44. 320-3. 10.1016/j.amepre.2012.11.008

[6] Lijuan Marissa Zhou, Cathal Gurrin, and Zhengwei Qiu, "ZhiWo: activity tagging and recognition system for personal lifelogs," 2013 In Proceedings of the 3rd ACM conference on International conference on multimedia retrieval 2013, Association for Computing Machinery, New York, NY, USA, pp. 321–322

[7] Duane, Aaron, Rashmi Gupta, Liting Zhou and Cathal Gurrin, "Visual Insights from Personal Lifelogs: Insight at the NTCIR-12 Lifelog LIT Task." *NTCIR* (2016).

[8] Melanie Swan, "Emerging Patient-Driven Health Care Models: An Examination of Health Social Networks, Consumer Personalized Medicine and Quantified Self-Medicine and Quantified Self-Tracking," International Journal of Environmental Research and Public Health 2009, Volume 6, Issue 2, pp. 492-525

[9] Yi Chen and Gareth J. F. Jones, "Augmenting human memory using personal lifelogs," In Proceedings of the 1st Augmented Human International Conference (AH '10), Association for Computing Machinery, New York, NY, USA, Article 24, pp. 1–9

[10] Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz, Arthur A. Stone, "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method," Science, Volume 306, Issue 5702, 2004, pp. 1776-1780

[11] Svetnik, Vladimir, Andy Liaw, Christopher Tong, J. Christopher Culberson, Robert P. Sheridan and Bradley P. Feuston, "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling," *Journal of chemical information and computer sciences*, 2003, Volume 43, pp.: 1947-58

[12] Kumar, Manish and Thenmozhi, M., Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest. Indian Institute of Capital Markets 9th Capital Markets Conference Paper

[13] Yoav Freund and Robert E. Schapire. 1996. Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on International Conference on Machine Learning (ICML'96). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 148–156.

[14] Jerome Friedman, Trevor Hastie and Robert Tibshirani "Additive Logistic Regression: A Statistical View of Boosting", *The Annals of Statistics* Vol. 28, No. 2 (Apr., 2000), pp. 337-374

[15] Panagiotis Sentas, Lefteris Angelis, "Categorical missing data imputation for software cost estimation by multinomial logistic regression", Journal of Systems and Software, Volume 9, Issue 3,March 2006, Pages 404-414

.

Source Code: ***https://github.com/jpete17/Lifelogging-Human-Activity-Classification***