Human Behavior Prediction and Analysis Using Machine Learning-A Review

Monali Gulhane¹, T.Sajana²
¹Student, K L University, Vijayawada.
²Professor, K L University, Vijayawada

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: Nowadays many trends are being in the area of medicine to predict the human behaviour and analysis of patient behaviour is being studied but the technical difficulty of cost efficient method to predict the behaviour of user is overcome in the proposed researched methodology. The mental health of the used can lead to good immunity system to be healthy in this pandemic of COVID-19. Hence After a detailed study on different human health disease classification techniques it is found that machine learning techniques are reliable for the feature extraction and analysis of the different human parameters. CNN is the most optimum choice of classification of diseases. Feature extraction and feature selection is automatically managed by the CNN layers, which reduces the training speed. Techniques like sensor-based feature extraction like EEG, ECG, etc. will be further explored using machine learning algorithms for detection of early detections of diseases from human behavior on different platforms in this research. Social behavior and eating habits play a vital role in disease detection. A system that combines such a wide variety of features with effective classification techniques at each stage is needed. The research in this paper contributes the review of the human behavior analysis through different body parameters, food habits and social media influences with social behavior of the person. The main objective of research is to analysis theses different area parameters to predict the early signs of the diseases.

Keywords: Machine Learning, CNN, EEG, ECG, Social Behaviour.

1. Introduction

Simulation of human behavior is a major aspect of implementation areas such as health care and research scientific studies social behaviour. Besides precise estimate, it is also essential to the ability to recognize the responsibilities of factor in human behavior as well as provide hypotheses for the predicted behaviours. Providing one such wider and broader credibility in the structures and the likelihood of actually adopting the systems, thus driving engagement and loyalty. Many other predictive models, however, do not offer an explanation for the behaviours they anticipate for early prediction of diseases. In the this paper, we are studying the research problem, prediction of human behaviour with justifications, for health care treatment systems, for health related to influence of social network(online).The research is based on different parameters like body parameter(heart, brain depending upon the datasets),social media influences and search history data, food habits of the person. Social science hypotheses about social cognition can notify our insight into how people humans express social media behaviour The review of these parameter is the objective of our paper.

2. Literature Review

2.1 Analysis of Human Body Parameter to Study the Behaviour.

The application range of recognition of human behaviour has been expanding with the rapid development of smart devices. Human behaviour recognition based on deep learning has become a new direction of research because of the problems of low efficiency and low accuracy of traditional artificial feature extraction behaviour recognition[9] the author studies the images using the conventional neural network to analysis the behaviour. This paper proposes an approach collaborative arrangements-AFRRNet model for recalibration of features. This article prepared the CNN model, RNN model, and AFRRNet model on the updates automatically data set, SAImage Net, to measure the accuracy of the AFRRNet model. This paper proposes a quaternion spatio-temporal convolutional neural network coherent with human behavioural understanding

Several theories related to human behaviour are under research related to different Yulia Tyshchuk et.all [1] in work The need for an integrative, thorough conceptual model of human behavior as expressed in social media. The article identifies the whole shortcoming as it relates to behaviors that change in response to notable events.In this work Three behaviors linked to warning response process was modelled: 1) the warning was obtained and propagated; (2) seek further information / confirmation; and (3) take the mandatory action

The feature under research stated by the author [1] states the use of natural language processing, along with sentiment, modality, polarity, emotion and arousal

2) social network analysis for community detection.

Personalized behavioural / perceptual prediction is extremely being applied using the Machine Learning (ML) regression challenges. The precise ML regression algorithm and sampling technique are two significant factors
that affect predictive accuracy in a non-trivial way. Nevertheless, the impacts of the ML regression algorithm and confidence interval on individualized performance of the behavioural / cognitive prediction were not assessed[10] Zaixu Cui et.al has tried to overcome this issue by using six different generally used ML regression algorithms such as 1. ordinary minimum square (OLS) regression, 2. least absolute shrinkage, 3. selection operator (LASSO) regression, 4. ridge regression, 5. elastic-net regression, 6. linear support vector regression (LSVR) and pertinence vector regression (RVR) to feature the specific behavioural/cognitive predictions based on various sizes of sampling, but in spite of using six algorithms the research is having drawback of accuracy prediction in the behaviour. The study state the complex sample size of nueral information of the brain. Hence in our research to overcome the accuracy issue we have tried to implement the hybrid combination of body parameter and human behaviour analysis through the dataset of EEG and ECG.

Shuqin Wang et.al [11] in Dynamic Human Behavior Pattern Detection and Classification Detection and classification of human dynamic behaviour based on video plays an important role in understanding human movement, and it has many interesting factors Crime watch applications, security surveillance, sports, and so on.

This paper stated that the initial investigation into the detection and classification of dynamic human behaviour patterns based on real-time video data. Researchers stated that the study and compare future work Various time-based Machine Learning Models Study by experiment.

Seema F. Al Raisi et.al[12] states that analysis of human observer behaviour in images and videos. Practical inspection in many areas is carried out using data captured through eye tracking apparatus. Author analyzes such analysed using different approaches which creates valuable knowledge and capabilities decision making on the material. The author has researched to investigate the attention of the human observer to different parts of the human body, area of head, area shirt, pants/legs and any other object which is carried or pushed or pulled, etc. Human body area containing an object outside the human body the Tobii Studio software DAOI tool has been used to create dynamic areas of interest

There are different factors based on which human behaviour can be predicted such as blood pressure such research is observed in the paper by Po-Han Chiang et.al in Personalized Effect of Health Behavior on Blood Pressure: Machine Learning Based Prediction and Recommendation [13] the author proposed that Author suggested using Machine Learning (ML) methods for building a custom model to forecast daily BP using the cultural BP and health behaviour of an individual; and estimate the impact of healthy behaviours on individuals BP / BP . In this research author explore the daily BP prediction and its relationship between wearable data collected on health behaviour. In addition to numerical prediction of daily BP, we propose to investigate the personal effects of health behaviour on daily BP with the relative importance generated by the proposed ML model among different predictor variables (features). For implementing the above features author has used machine learning model using Random Forest (RF) RF with Feature Selection (RFFS), which performs RF-based feature selection to filter out unnecessary features

2.2 Human Behaviour Analysis for detecting early signs of diabetes

In Ioannis Kavakiotis et.al [2] states in his research author reviewed that study to carried out a systematic review of the application areas of machine learning, data mining technologies and tools in the field of disease research with resp ect to (a) diagnosis and prediction, (b) diabetes mellitus, (c) genetic heritage and environment, and (e) first category health care and management which appear to be the most popular. The study of this review by us observed that there is complete lack of data concerning a) lifestyle and behaviour, b) inheritance, and c) linkage with other path physiological conditions e.g. Alzheimer’s disease. Therefore we tried to carry the review further for the disease analysis of such limitations. In the study Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes [3] we observed that machine learning algorithm can be used to develop a compact guide regarding modelling options and strategies of machine learning and a hybrid system focusing on the prediction of BG dynamics in type 1 diabetes. These study can be one of the reliable feature for detection of diseases at the early stages using machine learning algorithm. The paper states that Diabetes, also known as diabetes mellitus (DM), is a collection of metabolic disorders cause by the identification of high blood glucose levels over extended length of time. The early prediction signs are incorporate urinations, increase in hunger and thirty .If this diseases not treated on time can cause the life risk to the patient. Hence author has introduced the to implement a system on prediction of the blood glucose level using machine learning algorithms . The principal limitation of this research is the absence of a

1. To estimate the intake of carbohydrates, which is mainly done manually by individual users and is susceptible to errors that can seriously affect predictive performance.

2. The lack of a universal approach for
estimating and quantifying the approximate effect of physical activity, stress and the incidence of BG infections.

2.3 Human Behaviour Analysis for detecting early signs of Cancer

The machine learning algorithm is also used for prediction or study the early signs of many diseases one the major disease world is facing in this situation of the pandemic COVID-19 is mental illness or Major Depression Disorder (MDD) is a psychological condition defined as a depressive mood state with an issue duration / severities (minimum two weeks). Depression has an adverse effect on one's daily life, causing individuals to become unwilling or unable to carry out activities. In the recent study by Siyang Song et.al [4] proposed a system called a one-hand system to automatically identify primitive human behaviors such as Gaze Directions, Facial Action Units (AU), etc. as low-dimensional multi-channel time series analysis, which can then be used to create two sequence descriptors. The main area of the research in [4] is to explain how to use the sensory behaviour detected dynamically in low dimensions, descriptive tags for raw video data, instead of high dimensions, assessment on depression; 2. how to retrieve dynamics on a lengthy period feature of variable length data from time series.

The human behaviour analysis can also be analysed to predict the globally dangerous diseases like cancer. Although there is vast research done under the symposium of the detection for cancer or analysing the early signs of cancer using machine learning algorithm. It’s impossible to cover the relevant research on prediction or early signs of cancer, the author in [5] A Machine Learning Ensemble Classifier for Early Prediction of Diabetic Retinopathy states that the statistical methods discussed here are based on various supervised ML technology and different input and feature vectors. In view of the increasing pattern in the application of ML algorithms in cancer research, researchers are presenting here that the latest publications that use these techniques as a model for cancer risk or patient outcomes.

2.4 Human Behaviour Analysis for detecting early signs of Heart Diseases

The human behaviour analysis in machine learning is increasing its usage day by day with the advent of new research and technologies. Our research focused on the analysing the human behaviour then prediction of the early signs of diseases can be featured. In the proposed system we consider to predict the early signs of Alzheimer’s, Cancer and Diabetes or heart attacks the hybrid algorithm of the prediction of early signs of these diseases is the main objective of this review. The diseases can be featured by collecting the datasets of EEG and ECG from the doctors with the study fields like food habits, mental state(sleep orders), social media influences. There are many research working on these parameter but to overcome the few lacking we are trying to propose the hybrid group of parameter. The review is further done for the different parameter of prediction of cancer as stated in [6] by Rohit Rastogi et.al Scientific studies use machine learning (ML) algorithms to cause a person’s heart disease risk factors, depending on a number of medical history characteristics. The author has used different data mining (DM) techniques such as Naive Bayes, decision trees, support vector machines, and logistic regression which is used to analysed in the heart disease database. The accuracy level of all the techniques implemented of various algorithms is measured and the algorithms features were compared. The result of this experimental analysis is a 0 or 1 result that poses no danger or danger to the individual.

As stated, our proposed approach of hybrid parameters for detection of the early signs of diseases from the study of human body parameter or behaviour the research done explains that electrocardiogram (ECG) is a signal is used to measures the electric events or activity of the heart. Far Ibrahim Alas an and Mamoon Younes [7] have explain the use of ECG signal for the detection of heart diseases moreover the study in future tend to imply the classification that can be used with other types of datasets such as stress and clinical datasets. Machine learning uses artificial intelligence, and is used in data analytics to solve many problems. One is prevalent machine learning application is outcome prediction based on existing data. The Algorithm learn patterns from the existing dataset and then use them to predict the unknown dataset result.

3. Analysis of Parameters for Human Behavioural Prediction

The analysis for the studying the human behavioural parameter leads to the prediction of early signs of the disease. In this section we describe various methods that have been proposed by researchers over the years to predict human behaviour parameters for different kinds of scenarios. The work in [14] proposed by Peter G. Fennell and others, presents a non-bias approach to model human behaviour patterns. In their approach, structured sum-of-squares decomposition or S3D algorithm is proposed. This algorithm can take into consideration any kind of human behaviour pattern and then identify weights and relations between them. These weights are then converted into smaller homogenous blocks that can tell about a person’s behaviour individually. For instance, the algorithm can cluster data based on tweets and e-commerce patterns, to identify
the proclivity of a user towards buying a certain brand of medicine. Their algorithm is based on decision trees, and therefore is limited by the speed and performance capabilities of the decision tree algorithm. This work can further utilize pattern matching with the help of deep learning algorithms to enhance its efficiency. They have used the disease types like Breast Cancer & Parkinsons for analysis, but the concept can be extended to any disease related to human behaviour in general. This work can also take reference from the work done in [15], wherein researchers Ashish Patel and Jigarkumar Shah have proposed a simplistic but dynamic approach to behavioural pattern prediction. Based on their research, the human behaviour is generally unpredictable, and prediction algorithms must always take into consideration the ever-changing models of human behaviour. To this end, they have proposed the solution under ambient assisted living (AAL) conditions for elderly people, but it can be extended to any age group in general. Their approach uses visual analysis, wherein the human postures are studied and then are predicted into normal and abnormal using deep learning algorithms. They also propose the use of a feedback mechanism, that can further help in improving the prediction performance.

Social networks have become a crucial part of our day-to-day lives. The work in [16] analyses various aspects of online social networks, and presents a model which can be utilized for human behavioural analysis. They have proposed Social Influence Deep Learning or SIDL framework in order to do so. Using SIDL deep learning methods like CNN and Deep CNN can be combined with data from different social networks like Plancast, Foursquare, Facebook, Twitter and others. This combinatorial data analysis is helpful in assisting the system about various human behavioural patterns. For instance, tweet data can be combined with Foursquare location data to identify the user’s mood and the location of the mood. Based on this information, if the user is standing on a bridge and is showing suicidal patterns, then emergency services can be informed and the user can be rescued. Such data can be directly linked with the user’s health patterns. More and more depressive social media data in the form of post updates, comments, group information, music preferences, and others can be combined together to identify long term and short-term diseases in the user. The SIDL framework does this task with more than 85% testing set accuracy, which can be a very good research milestone in this field of study. Sensors can also be a major source of behaviour prediction data. The work in [17] showcases that passive monitoring using sensors like fall sensors, medical devices and other equipment can be useful when predicting user’s behaviour. Sensor data from a sensor carpet is combined with pressure sensor data, motion sensor data, and other medical sensor data like temperature, heart rate, BP and stress levels to form a data vector. This data vector is combined with the end user’s actual behaviour to form a training dataset. The training dataset is given to a linear classifier to evaluate the behavioural pattern of the user. This information can be further researched, and techniques like convolutional neural networks can be used in order to perform complex classification operations that combine diseases from different actions of the users. Such a mechanism is devised in [18], wherein data similar to [17] is processed, but the behavioural patterns used are majorly related to patient’s health. The researchers are using higher order classifiers like SVM, and fuzzy logic to perform the task of disease identification. The classifier is combined with a boosting unit to perform hypertension classification. It is proven that SVM and Naïve Bayes can outperform other linear classifiers. But, the performance of CNN is not evaluated, and thus can be a research area to be explored.

Smartphone usage can be another parameter for prediction of user behaviour. The work in [19] uses text messages and phone call patterns to evaluate the behavioural patterns of the users. If a user is trying to communicate constantly with someone, then it might either relate to urgency or anxiety. But, if the user is trying to contact someone continuously for a greater number of days, without getting a response from their side, then it might indicate anxiety or lack of attention. All these issues are indicative of abnormal behaviour, and can be used to identify different disease classes. The work in [19] claims that they are able to achieve an accuracy of more than 90% with the help of such analysis with the help of DBSCAN classification. It is recommended that such an innovative research can be further investigated with the help of CNN algorithms, and more secondary disease patterns can be evaluated. Moreover, this can be combined with consecutive behaviours in order to further improve the system performance, and add more disease types to it. For instance, the work in [20] is used for prediction of fraudulent behaviour using tele-communication datasets. This work uses higher order classifiers like LSTM (long-short term memory) in order to perform classification of user’s telecom data into one of many categories. Their model is named as Network-enhanced History Attention-based-LSTM, because they are using network history such as call logs, messages, location of call, and other parameters for behaviour prediction. Such information can also be another source of behaviour pattern prediction for user’s health monitoring. For instance, data related to type of text messages being sent, can be an indicative of the person’s mental health, while the data about the person’s call and location details can give a rough estimate about the person’s thought process and movement patterns. These movement patterns can be used to trace if the person is going to certain areas frequently, which might further indicate the source of diseases or infections the person might catch after visiting certain areas. Such secondary parametric analysis seems like a long shot, but can easily be used to conclude about a person’s health status.
It is known that a person in stress is bound to do things in rage [21]. The work in [21] uses this property, to identify the mental state of the user by their keyboard typing patterns. They have used smartphone keyboards and physical keyboards in order to find out if the user is in stress or not. This can be a very good feature for identification of emotions like rage, stress, sadness and excitement. Using the sensors like input from keyboard, gyroscope & accelerometer, and then combining them with linear classifiers is able to identify the stress levels with up to 87% accuracy. This accuracy can further be improved with the help of complex classifiers like SVM and CNN. Emotions like Tired, Happy, Stressed, Energetic, Angry & Interested were identified using this research. It is recommended that this research be extended by linking these emotions with health-related factors. As mentioned in [16] social behaviour plays a very important role in predicting user behaviour. The work in [22] extends this concept by taking election user tweet data, and then combines it with deep learning-based classifier to evaluate user behaviour patterns. These patterns are more focussed towards election scenarios, but by changing the dataset, the same concept can be applied to any kind of human behaviour. The researchers have used deep learning CNN Models to achieve an accuracy of more than 95% to identify user’s behaviour, and thus it can be used as an accurate tool for performing the task of effective behaviour prediction.

Diseases and health issues are more prevalent in older people. The research in [23] addresses this issue by gathering data from smart devices connected with older people to identify their behaviour patterns. In order to perform this task, they have proposed an “Elderly people by room activities description language” that uses simple “what”, “who” and “when” questions to identify the context of the behavioural event, and then accordingly classify it into different categories. This information is combined with a Temporal Hierarchical Hidden Markov Model to identify behavioural patterns of Elderly people. An interesting research is done in [24], wherein data of newly wed couples is taken in order to identify behavioural patterns that account for marriage decisions. One of the conclusions from their research indicates that a neurotic husband usually gets wed to a lesser extrovert and open wife. Such interesting research articles open up newer features that can be used in order to facilitate the accuracy of human behaviour prediction. Based on this research, it is possible to evaluate patterns of current marriages, and conclude whether the mental state of both the partners might be good or not, and then link these results with personality traits to evaluate the mental condition of the users. This condition can be an indicative of long-term disease prediction for the users. A similar work is done in [25], wherein group-based analysis is done for prediction rather than a dual-person analysis. Group behaviour patterns can easily help in identification of user’s mental and social habits. It is said that a person is a product of the group she/he is a part of. This majorly relates to the person’s mental and health conditions. A group with healthy individuals will always have a less chance of health disease than a group with obese and unhealthy individuals. Thus, group behaviour can also be a very important part of analysing the individual behaviour of a person. The work in [25] is able to achieve more than 90% accuracy in this domain, and thus can be taken as a reference for any kind of group behaviour analysis system.

Extending the work in [19], the work in [26] takes into consideration smart-phone data like location, communication, media consumption, communication and social behaviour, music consumption, app usage, mobility, overall phone activity, and day- and night-time activity are distinctively used in order to predict the user’s behaviour. The system is able to achieve more than 80% accuracy, and thus can be used as a feature vector while defining user patterns. This work can be extended with the help of the work in [27], wherein visual features are combined with behavioural features in order to identify final user behaviour. These features are classified with the help of CNN-based classifiers to obtain an accuracy of more than 90%.

There is always a connection between human health and social networks. But the work in [28] goes further and identifies patterns in health social networks using social restricted Boltzmann machine. Their work combines self-motivation, implicit and explicit social influences, and environmental events together in order to form an integrated behavioural prediction model. The features are taken from the health social network YesiWell. These features include, but are not limited to BMI, fitness, cholesterol, the number of off-line events in which each user participates, individual sending and receiving messages, the number of goals set and achieved, Wellness score, the number of competitions each user joins, the number of goals set, the number of exercise days, walking/running steps, the distances, speed walked/run, the increase of BMI slope and/or decrease of Wellness score and others. All these features form the basis of this research, and can be used to identify diseases with a staggering 90% accuracy. Another interesting work is done in [29], wherein social lifestyles are predicted with the help of user movement patterns. These movement patterns are taken from user’s smartphones, cell signals, and other positional data. The system also uses places of interest (PoI) in order to further facilitate the behaviour classification. This information when given to a CNN classifier, can produce accuracy levels of more than 90%. This work is further extended in [30], wherein the next-location of the user is predicted using data similar to [29]. This prediction is also done with the help of statistical data like [29], and is able to predict the user’s next location with more than 90% accuracy. From these research studies it is evident that there are multiple features which can be used in order to find out the human behaviour patterns. These features range from generic social media data, to more specific smart phone usage patterns of the user. Moreover, group-based analysis can also be a helpful tool in identification of human behaviour, which can further be linked with health conditions of the user under study. Classifiers like
support vector machine and convolutional neural networks outperform other methods for identification of behavioural patterns.

4. Conclusion and Future work
From the analysis we can identify that diseases like diabetes, cancer, heart diseases, depression, mental disorders and others can be linked with human behaviour. This behaviour is linked with the daily habits of the user including food habits, social media patterns, health parameters, telecom data, the person/people with whom the user usually spends their day with, smartphone data, and others. All these data patterns can be used in order to identify a user’s behaviour, and thereby predict the user’s health status. Moreover, these parameters can be mixed and matched in order to improve the accuracy of behaviour prediction. In this text, it is observed that any kind of data which can be observed from a person’s point of view can be used to analyse the person’s behaviour, and thereby is useful for prediction. Classification algorithms like CNN, SVM, HMM and Deep nets can be very useful in order to perform highly accurate behavioural prediction. It is recommended that researchers must use a combination of behavioural patterns in order to evaluate the best possible feature combinations for identification of human behaviour.

References