USING DEEP LEARNING PREDICTIONS OF SMOKERS' BEHAVIOUR TO DEVELOP A SMART SMOKING-CESSATION APP

M A Abo-Tabik

PhD 2021

USING DEEP LEARNING PREDICTIONS OF SMOKERS' BEHAVIOUR TO DEVELOP A SMART SMOKING-CESSATION APP

Maryam Abdulmunem Salih Abo-Tabik

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF
MANCHESTER METROPOLITAN UNIVERSITY
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Department of Computing and Mathematics
Manchester Metropolitan University

2021

Contents

List of Tables

List of Figures

Abstract

The number of new smoking-cessation apps had increased in recent years. Although these offer accessible and low-cost support to smokers, they often lack scientific understanding of nicotine addiction, and rely on smokers' self-reporting their cravings / environmental factors; a method widely acknowledged to be unreliable. This PhD presents two novel deep-learning models for automatic smoking events prediction. Both models combine machine-learning with Control Theory Model of Smoking (CTMoS), to enable the prediction of smoking events based on both internal (nicotine level) and external (e.g. location) factors. This offers a way to overcome limitations of previous apps.

The first model, combined CTMoS with a 1D Convolutional Neural Network, using raw accelerometer and GPS coordinates as input. Result indicated good prediction of internal craving factors (e.g. nicotine level and craving); but smoking events prediction required improvement, as the f1-score were 0.06, 0.14, 0.24, and 0.4 for predicting a smoking event 5, 15, 30, and 60 -min (respectively) prior to its occurrence.

The second model combined 1D Convolutional Neural Network with the Bidirectional Long Short-Term Memory method, to create a deep learning model with Genetic Algorithm for hyperparameter selection. The model used the same 3- accelerometer values as input, but the 3-GPS coordinates were replaced with coded location data (five most smoked locations). These changes improved smoking events prediction with average f1-score of 0.32, 0.59, 0.71, and 0.8 for predicting a smoking event 5, 15, 30, and 60 -min (respectively) prior to its occurrence.

This PhD achieved its three goals: minimize user input (by using data collected from phone sensors); improve scientific understanding of factors that influence smokers' behaviour (by evaluating the relative contribution of different factors), and developing a state-of-the-art model that enables the automatic prediction of smoking events. As such, outcomes of this PhD lay the foundation for future development of smart and personalized apps that can provide real-time personalized support for smokers.

Declaration

No part of this project has been submitted in support of an application for any other degree or qualification at this or any other institute of learning. Apart from those parts of the project containing citations to the work of others, this project is my own unaided work. This work has been carried out in accordance with the Manchester Metropolitan University research ethics procedures, and has received ethical approval number 0441 and 24517.

Signed: M A Abo-Tabik

Date: 07/ 01/ 2021

Acknowledgements

I would like to thank my supervisors Dr. Nicholas Costen, Dr. Yael Benn and Dr. John Darby, for believing in my work, without them this thesis would have not been possible. I am especially thankful for them supporting me immensely on many levels. It was a great pleasure working with them.

I also would like also like to thank Naughton et al. (2016) the providers of the QSense dataset.

I want to thank my family especially my parents for their unconditional support, without them I would have never been able to reach this far.

I dedicate this thesis to Faisal.

Abbreviations

1D-CNN	1Dimensional Convolutional Neural Network
ANN	Artificial Neural Network
AUC	Area Under a Curve
BiLSTM	Bidirectional Long Short-Term Memory
CNN	Convolutional Neural Network
CTMoS	Control Theory Model of Smoking
DNN	Deep Neural Network
DT	Decision tree
EMA	Ecological Momentary Assessment
GA	Genetic Algorithm
GPS	Global Positioning System
LSTM	Long Short-Term Memory
mHealth	mobile Health
ML	Machine Learning
MLP	Multilayer Perceptron
MMU	Manchester Metropolitan University
MSE	Mean Square Error
NRMSE	Normalised Root Mean Square Error
OOB	Out-Of-Bag
ReLU	Rectied Linear Unit
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
UI	User Interface

Chapter 1

Introduction

Smoking is considered one of the leading causes of death in the UK (ONS, 2018), and increasing awareness about the harmful impact of smoking (e.g. health, expenses, etc.) makes many people consider quitting smoking. A report published in 2018 stated that in developed countries, around 55% of smokers consider quitting; it also states that, despite the availability of a wide range of treatments for nicotine addiction, less than 1 in 10 smokers successfully manages to quit smoking (Babb et al., 2017; Creamer et al., 2019).

It was found that there is a strong relation between smoking craving and smoking events (Shiffman et al., 2009), such that craving is considered to be the main reason for relapse (Dunbar et al., 2010). According to the NHS (2018) website,

> "Cravings happen because your body misses its regular hits of nicotine. There are 2 types of craving. The steady and constant background craving for a cigarette decreases in intensity over several weeks after quitting. Sudden bursts of an intense desire or urge to smoke are often triggered by a cue, such as having a few drinks, feeling very happy or sad, having an argument, feeling stressed, or even having a cup of coffee. These urges to smoke tend to get less frequent over time, but their intensity can remain strong even after many months of quitting.".

In other words, the urge to smoke or smoking-craving, can be linked to internal factors; such as the need to restore the nicotine level in the blood or the smoker's mood. Maintaining nicotine level will prevent withdrawal symptoms (Stevenson et al., 2017). For example, most smokers consume cigarettes more often in the first two hours

of the day after waking up than they do in the rest of the day. Mostly, this is due to the need to refresh the nicotine level after a long night's sleep (Chandra et al., 2011).

However, internal factors including nicotine level are not the only triggers to smoke. Shiffman et al. (2014) collected data from 194 participants who reported each cigarette they smoked and its related situation. Results revealed an association between smoking and different external factors that are related to the smoker's personal preferences, such as the locations where they mostly smoke in (e.g. home, bar, etc.) or the type of concurrent activity, for example eating, drinking, or for some smokers working. For example, many smokers reported that they do not feel the urge to smoke when they work even when the work refers to simple home chores.

Nevertheless, internal factors, and particularly nicotine level, have the strongest influence on smoking, in which it creates what it looks like a daily smoking pattern. Over time the dependence on nicotine will increase (Chandra et al., 2007; Mathieu-Kia et al., 2002), and the patterns start to differ between smokers based on external factors (Grainge et al., 2009; Hughes et al., 2017; Perkins et al., 2013).

It can therefore be concluded that both internal and external factors influence smoking craving, and in turn, smoking patterns. This makes it hard for smokers to quit as there are potentially many diverse triggers that can lead to relapse. In particular, once a person gives up smoking, even if they use nicotine replacement, it is hard to isolate a person from their environment or change their daily routines. As a result, smokers are likely to continue to experience cravings based on environmental cues, making it hard for them to quit.

However, with advances in mobile technology and the increasing use of smartphones (Holst, 2019) it is possible to provide mobile services that give timely and targeted interventions to support smokers through the difficult moments. This can be delivered in the form of support text messages (e.g., "You have done well to abstain from smoking for a month!") or suggestions (e.g., "You may be tempted to smoke soon, maybe take an exercise break."), or other (e.g., image of the smokers aged face if they continue smoking) to targets critical moments when the smoker mostly feels the need to have a cigarette. This is likely to be effective as it may distract the attention of the smoker from their craving and relieve the stress associated with it. Such a service will interrupt the normal flow of the environment, and distract the smoker attention from the present external craving factor.

This type of smartphone health services has been referred to in recent years as Mobile Health (mHealth) (Mobasheri et al., 2014). Among other things mHealth includes

smartphone apps that can deliver remote health services (e.g. remote monitoring, data collection, remote medical support and intervention, etc.) at a low cost. As this technology can be reached by people wherever they are, it is considered to be particularly useful for reaching groups that have been traditionally categorised as hard to reach (McClure et al., 2016).

mHealth has been shown to provide effective interventions in several fields including pain management (Rosser and Eccleston, 2011), insomnia (Pulantara et al., 2018; Walsh and Groarke, 2019), eating disorders (Chen et al., 2014; Tregarthen et al., 2015), stress management (Ferdous et al., 2015), and smoking (Businelle et al., 2016; Naughton et al., 2016).

Smoking cessation apps, like many other mHealth apps, are smartphone software that can support smokers in the process of quitting smoking and assist them in tracking their progress. This support can include advice on changes to lifestyle routines, rewards for passing quitting milestones, support for the user in resisting a smoking craving event alongside other methods. Importantly, these apps hold the potential of aiding smokers to quit due to their ability to provide remote support at a low cost (Ubhi et al., 2016).

However, while the smoking cessation app market has seen growth, only a small number of the apps have been scientifically examined or supported with randomised controlled trials (Abroms et al., 2013; Haskins et al., 2017). Furthermore, current smoking cessation apps do not make full advantage of advances in mobile technology (McClure et al., 2016; Vilardaga et al., 2019). An early approach to designing smoking cessation apps relied entirely on the user logging their progress on the app and reporting their smoking relapses. Despite this simple approach, these apps show promising results on the efficiency of using mobile apps to support smokers during the quit process (Ubhi et al., 2015).

A more advanced approach has the app delivering intervention messages at either periodic or random times. These messages have sometimes been personalised to the user, which usually requires that the user completes a demographic form and report preferences and smoking history that are then used to select intervention messages. Examples include using the name of the user in the message, or suggesting healthier life routines if the smoker had reported this as a motivation for quitting smoking (Baskerville et al., 2015; Buller et al., 2014). Such personalised apps revealed better user engagement than the previous approach (Paige et al., 2018; Regmi et al., 2017). Most of the available smoking cessation apps provide such basic tools (e.g. calendars,

calculators, distractors)(Hoeppner et al., 2016).

While these apps are more efficient, they are still unable to take into consideration both the internal and external factors that influence cravings. As such, interventions may miss key moments when the smoker is at high risk of relapse. Moreover, most of the available apps use self-reporting as a method to report craving factors or relapse events, which was previously shown to be unreliable (Businelle et al., 2016; Webb et al., 2020). Furthermore, long-term self-reporting is more likely to be influenced by the 'Ostrich problem' by which people avoid reporting their behaviour, as it may be tiresome, unpleasant, or they may not be entirely committed to changing their behaviour (Webb et al., 2013). Therefore, automated data collection from mobile sensors can reduce the reliance on self-reports, and increase the reliability of the smoking cessation apps (Naughton et al., 2016).

To address this, advances in smartphone technology can be employed to understand the smoker's daily pattern using built-in sensors such as Bluetooth (to detect the presence of other people) or Global Positioning System (GPS) (to detect high risk locations such as places the smokers used to smoke, tobacco sale point or in western culture, places where alcohol may be available such as pubs), and others. By using these technologies smoking cessation apps can minimise user input, while providing targeted and more personalised interventions, that can predict when the smoker is most in need for support, and deliver personalised and timely intervention messages.

This project addresses two important limitations that exist in previous work that attempt to assist smokers through the quitting process. The first, is the use of self-reporting of craving factors by smokers, and the second is the focus of smoking behav–iour-models on either internal or the external factors that influence smoking behaviour. This PhD project addresses these limitations by designing a system that feeds automatically collected environmental data from the sensors of smoker's mobile phone, into a deep learning model that combines representations of internal (mathematically modelled) and external (automatically collected) factors, in order to predict smoking behaviour in real-time. The resultant model sets the foundation for the design of smart, personalized state-of-the-art smoking cessation intervention apps, which rely entirely on automatically collected data and deliver intervention in real-time at low cost.

1.1 Project aim and objectives

Project aim: This project aims to design an automatic smoking event prediction model that can be employed in the development of smoking cessation apps. The model should be able to understand both internal and external factors that effects the smoking patterns; and also only use a passively collected data from smartphone device.

The main project objectives are:

- conduct literature review to highlight the gaps and limitations in the previously designed smoking cessation apps;
- validate the performance of Control Theory Model of Smoking (CTMoS) proposed by Bobashev et al. (2017) with human participants data;
- collect motion and environmental data from smartphone sensors, that incorporate real-time reported smoking events, motion and environmental data;
- develop a Deep Learning (DL) model that when combined with the CTMoS it can predict the individuals' smoking behaviour based on both internal and external factors.

1.2 Contributions

This project presents the state-of-art results on modelling smokers' behaviour and automated smoking event prediction, based on internal factors (i.e. nicotine level) and external factors (i.e. smoker's motion and location). This project main contributions are:

1. highlighting of gaps in the literature and the use of machine learning for smoking cessation intervention (published paper Abo-Tabik et al., 2021) (see chapter 2);
2. validation of the Control Theory Model (CTMoS, Bobashev et al., 2017) with data from human smokers, providing an update to the original model which used animal models Data for this validation was contributed by Naughton et al. (2016) (see chapter 4);
3. two models were presented in the thesis, in which they combine the Control Theory Model of Smoking (CTMoS) with Deep Learning (DL) approaches. Both models can be employed later in the development of the smoking cessation app,

to minimize the reliance on self-reporting approaches, and including both internal and external factors when predicting a smoker's behaviour:

- the first model combines 1D Convolutional Neural Network (1D-CNN model) with CTMoS to predicts smoker's behaviour base on external factors (smartphone built-in accelerometer sensor readings, and GPS coordinates) which feed to the 1D-CNN, and internal smoking factors (i.e. nicotine level, smoking craving, and withdrawal) which generated using CTMoS; (see chapter 5 and published paper Abo-Tabik et al., 2020);
- in the second model it presents an improvement in the performance of the first model, where it combines the 1D-CNN with Bidirectional Long Short-Term Memory (BiLSTM) models to be able to extract correlated patterns from the input vectors. In addition, the second model generalises the design of the first model by replacing GPS coordinates with coded location data; in which opens the potential for future development to make the app publicly available (see chapter 6).

1.3 Publications

- **Abo-Tabik**, M., Costen, N., Darby, J. and Benn, Y., (2019), August. Decision Tree Model of Smoking Behaviour. In 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld /SCALCOM /UIC /ATC /CBDCom /IOP /SCI) (pp. 1746-1753). IEEE.

- **Abo-Tabik**, M., Costen, N., Darby, J. and Benn, Y. (2020) 'Towards a smart smoking cessation app: a 1D-CNN model predicting smoking events.' Sensors, 20(4) p. 1099.

- **Abo-Tabik, M.**; Benn, Y.; Costen, N. (2021) 'Are Machine Learning Methods the Future for Smoking Cessation Apps?' Sensors 21, no. 13: 4254.

1.4 Thesis structure

- **Chapter 2**, this chapter gives a background on smoking behaviour modelling and the available types of smoking cessation apps.

- **Chapter 3**, this chapter gives a theoretical background on machine learning and deep neural networks methods that are going to be used in this project. It starts with a general machine learning background and the classical supervised models; support vector machine and decision tree. then the chapter goes in details explanation to 1D-CNN, and LSTM models and techniques to improve their performance.

- **Chapter 4**, this chapter describes the datasets used in this project in term of study design and analysis. Also, the chapter presents a second validation analysis to the Control Theory Model of Smoking (Bobashev et al., 2017) behaviour based on real smokers data.

- **Chapter 5**, this chapter describes the designed model of automated smoking behaviour prediction based on internal and external factors. The model combines the Control Theory Model of Smoking with 1D-CNN model.

- **Chapter 6**, this chapter describes the designed model of automated smoking behaviour prediction based on internal and external factors. The model combines the Control Theory Model of Smoking with 1D-CNN-BiLSTM model.

- **Chapter 7**, this chapter concludes the thesis with a summary of results and contributions and the potential avenues for future work.

Chapter 2

A design approach for smoking cessation apps

The chapter describes a range of models that have been used in the development of smoking cessation apps, proposes ways in which future development could be improved by making full use of the embedded features in smart mobile devices.

The chapter begins with a brief introduction to Opponent-Processes Theory of modelling addictive behaviour. Then it explained methods that can be used to understand the external factors that govern the smoker's behaviour.

The chapter highlights how the development of smoking cessation apps could be dramatically improved by using advanced ML methods, and how it can provide a better insight into patterns that govern smokers' behaviour, hence enable more personalised, better targeted and more timely interventions.

2.1 Opponent-Process theory for modelling internal smoking craving factors

Internal craving factors like nicotine level in the blood are difficult to collect regularly, and it requires special tools and/or a laboratory test to measure. However, science provides tools that can aid the understanding of what is happening inside the human body in the form of mathematical models. To this end, it has been previously been suggested that nicotine addiction, and in relation, nicotine level in the blood, could be modelled using a closed-loop control model. A closed-loop control model is a common instrumental technique. The basic form of the model uses a feedback signal from the

output, combining it with input to adjust the system parameters to maintain stability (Hughes, 2010). The general structure of a closed-loop control model is shown in Figure 2.1.

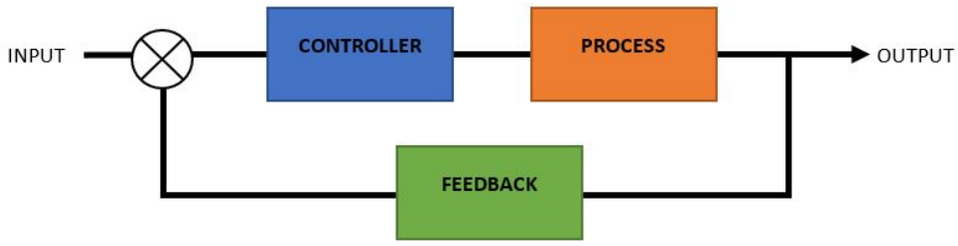


Figure 2.1: Schematic diagram for closed-loop control model.

A closed-loop control model has been shown to be successful in modelling behavioural actions such as eating and drinking, that are motivated by a need (such as hunger, thrust, etc.) to maintain stability (Fibla et al., 2010). Figure 2.2 shows drinking behaviour as a closed-loop model. The drinking rate $d\,(t)$ is derived as the difference between water volume $v\,(t)$ and the natural state V (i.e. when the organs have enough water) (Enquist and Ghirlanda, 2005),

$$d\,(t) = K\,(V - v\,(t)) \tag{2.1}$$

where K is a positive constant, that is different for each individual.

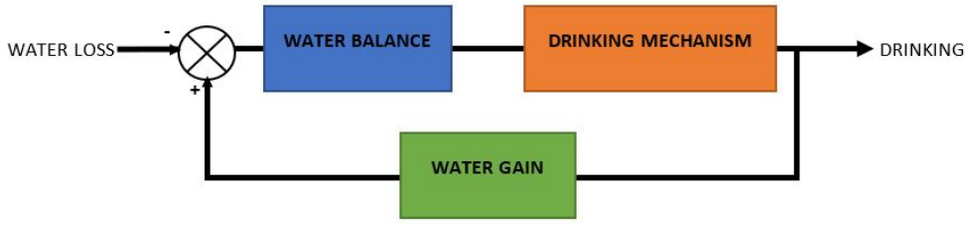


Figure 2.2: Schematic diagram of the drinking process as closed-loop control model, based on Enquist and Ghirlanda (2005).

However, modelling addictive behaviour as a closed-loop control model is a more challenging task, as it requires understanding the complexity of humans, and determining what elements should be included to accurately capture and model the addictive behaviour. Control theory offers the foundation for describing regulated systems that support modelling addictive behaviour (Gutkin and Ahmed, 2011).

Opponent process theory is a well-known method that is used to model a person's emotional state. According to this theory, emotions appear in opposite pairs, such that whenever a person experiences an emotion, its pair is suppressed; e.g., when someone is feeling happy their sadness is suppressed and vice versa (Metin and Sengor, 2012).

Solomon (1980) described addictive behaviour using the opponent-process theory. According to this model, following the first use of a drug, the user experiences pleasant feelings. However, over time the pleasant feeling decreases, causing the user to use the drug again. With repeated use, the user develops an addiction to the substance, accumulated withdrawal symptoms emerge, and the pleasant feelings decrease; this cycle results in the user increasing the frequency and/or the quantity of substance used in order to maintain intrabody stability. Due to the repetitive, rhythmical nature of addictive behaviours; control theory can offer the foundation for describing an opponent process theory that supports modelling smokers' behaviour (Ahmed et al., 2007) as shown in Figure 2.3.

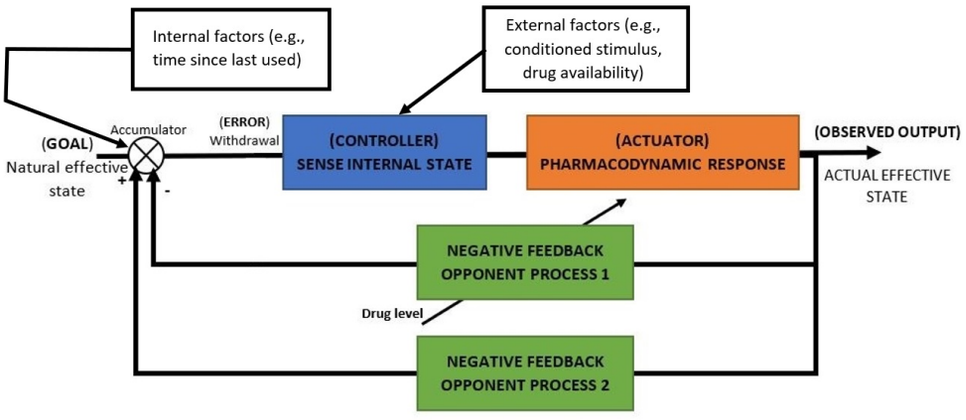


Figure 2.3: The opponent-process theory of addictive behaviour utilising control theory model
, based on Gutkin and Ahmed (2011).

In Figure 2.3, 'opponent process 1', which is the drug response, provides negative feedback and is usually modelled as a square wave because of its constant nature. 'Opponent process 2' is simultaneously initiated whenever 'opponent process 1' is initiated and gives positive feedback. The 'opponent process 2' signal is increased over time, which leads to a full or partial cancellation of the effect of 'opponent process 1' which leads to drug withdrawal syndrome (Gutkin and Ahmed, 2011).

Gutkin et al. (2006), proposed that nicotine addiction could be described using a neurocomputational model with an opponent process aspect. The model combines a

set of neural circuits at the molecular, cellular, and system levels, and accounts for several neurobiological and behavioural processes that can lead to nicotine addiction. This model is of particular interest for modelling smokers' behaviour because nicotine addiction is presented as a control system with a feedback loop, where the signals have different time scales ranging from seconds to weeks.

A recent study by Bobashev et al. (2017) utilised Control Theory with a formal multiscale opponent process representation to model smokers' behaviour. The resultant model was able to simulate internal craving factors (i.e. nicotine level, craving and withdrawal symptoms) in the long-term life of the smoker; the designed model is based on previous literature, and validated on data from an animal experiment. This model will be described later in more details in Chapter 4.

2.2 Using Ecological Momentary Assessment to provide target intervention

As mentioned in Chapter 1, when trying to understand smoking patterns it is important to consider both internal and external factors. While the model described above is useful for understanding the a substantial part of the internal processes (i.e., level of nicotine and withdrawal symptoms), external factors such as location or people around the smokers can not be modelled using the above mode.

An effective method for recording external indicator variables that are associated with smoking events is Ecological Momentary Assessment (EMA), a technique that enables the collection of data in real-time and in its natural environment (Shiffman et al., 1997). It has been used successfully to link smoking events with environmental information from smokers (Hébert et al., 2018; Koslovsky et al., 2018a). EMA has been employed in several smoking cessation apps, exploring the possibility of delivering interventions following the reporting of high risk factors for smoking such as alcohol consumption (Lynch et al., 2019), or stress level (Cambron et al., 2019).

EMA has been used by (Businelle et al., 2016) to detect the risk of smoking relapse; the study identified six common smoking craving factors and hence, a risk of relapse (urge to smoke, stress, recent alcohol consumption, interaction with someone smoking, cessation motivation, and cigarette availability). Participants in the study self-reported their activities, and this information automatically updated a diary located on their mobile phones. Results indicated reporting these predictors can improve the outcome of targeted intervention and prevent potential relapse.

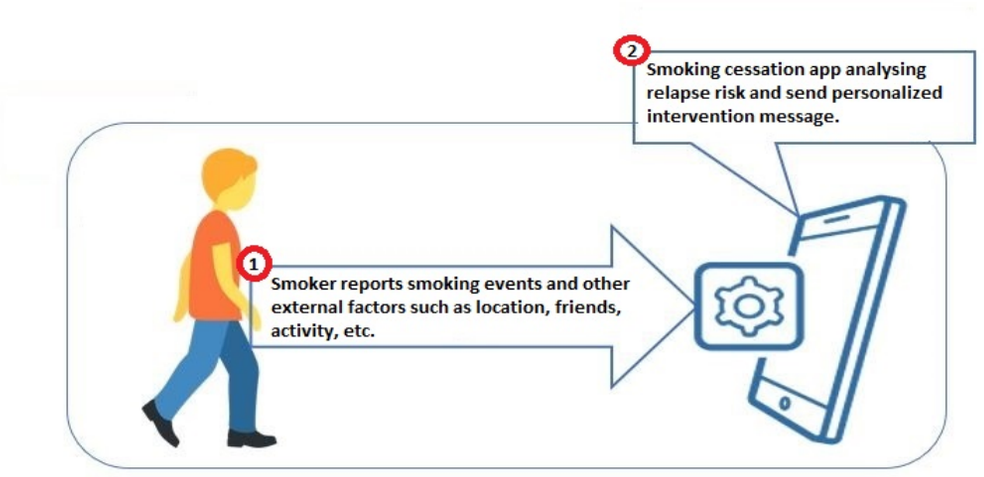
Location is an external craving factor that has been identified using EMA methods. Kirchner et al. (2013) was able to link location to craving level by collecting data from smokers who are trying to quit during the first month of the quitting process. The research also found that exposure to tobacco sales locations increases the risk of relapse.

Recently, EMA has been employed to improve smoking cessation apps. In these apps, the smoker receives motivational support or other support once they report a craving factor (Buller et al., 2014; Ghorai et al., 2013; Rodgers et al., 2005; Williams et al., 2018). Hébert et al. (2018) asked participants to complete an EMA diary 5 times a day for three weeks on their mobile phone app. The app then estimated the risk factor for lapsing in real-time and sent an intervention message that was either generic or tailored to the current level of risk and to self-reported resent of lap-risk. Results indicated that targeted messages were more effective than generic ones, re-enforcing the value of improving the timing and content of targeted messages.

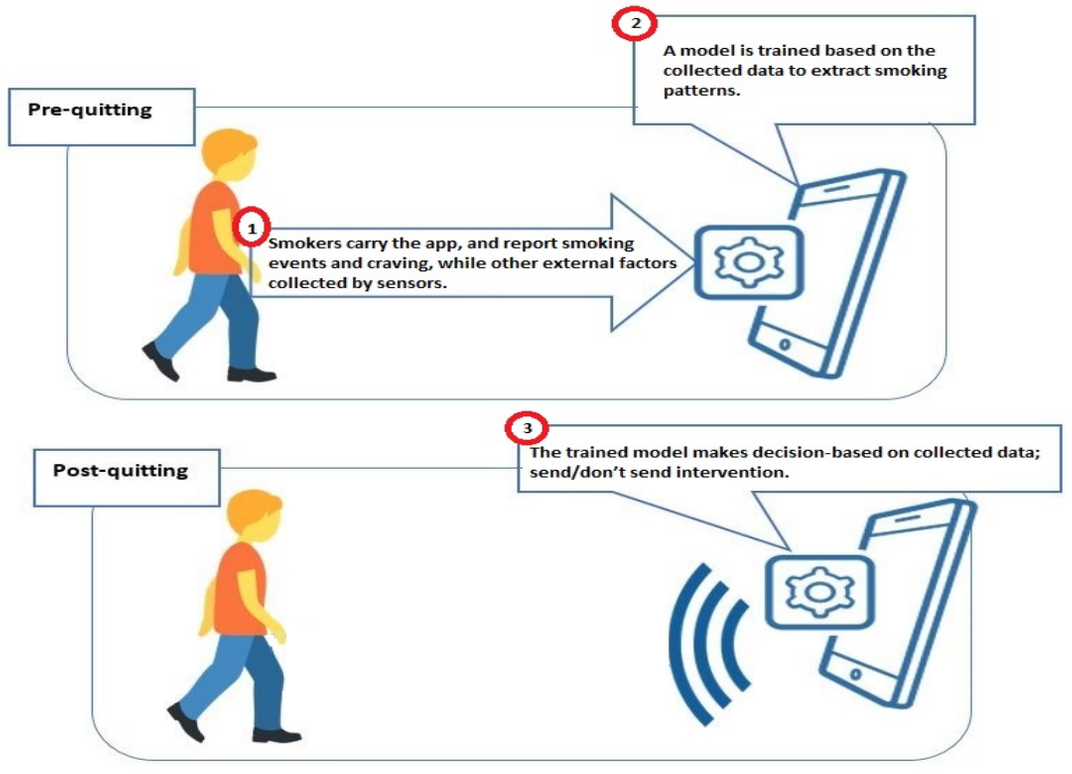
While the use of EMA improved the effectiveness of smoking cessation apps by making intervention messages more personalised, these apps still rely on self-reporting as a method to initiate an intervention. This is despite self-reporting previously shown to be unreliable (Gorber et al., 2009). However, advances in technology have enabled the collection of many of these self-reported variables automatically, avoiding the need to rely on user-input.

The main principle of automated intervention is to collect data from smokers during the pre-quit period. The collected data can then be used to understand smoking patterns, and the external (e.g. location) markers of the smoker's habits. Once the smoker quits smoking (post-quitting), the app can use these patterns to send interventions to the smoker in the most needed times without the need for initiating the request for intervention from the smoker. Figure 2.4 illustrates the difference between regular EMA apps 2.4(a) and automated EMA apps 2.4(b).

Naughton et al. (2016) developed the QSense app, which used combined self-reported and auto-collected EMA data to send automated triggered support message. QSense collected the data from 15 participants for two weeks prior to quitting day; each participant reported each cigarette they smoked to the app, and it used GPS service to detect and store the geo-location. If a participant reported smoking in the same location repeatedly, this location was saved as a relapse risk location. During the quitting phase, whenever the app detected that the participant came within a range of 3 metres near one of the smoking risk areas, an intervention message was sent. Results



(a) Regular EMA smoking cessation app.



(b) Auto-collected EMA smoking cessation app.

Figure 2.4: Approaches in designing smoking cessation apps.

from the study indicated that after the 28th day of smoking abstinent, only 50% of the participants reported smoking-relapse events, with an average of 3 lapses per participant. It was also reported that the first lapse usually happened after the 9th. The

participants of the study reported that target intervention and understanding the individual's smoking patterns was able to help to prevent some relapse moments, and that they generally felt positive about the experience.

While the Qsense automated intervention was effective, it can be improved by collecting more than one external factor. This in turn will require advanced analysis tools like Machine Learning (ML) in order to enable effective and efficient processing of the volume and complexity of the data.

2.3 Improving smoking cessation apps via Machine Learning : a literature review

Machine Learning (ML) provides a type of model that automatically improves its own performance through parameter tuning. The model uses a sample of data to learn how to perform a particular operation without being explicitly programmed or guided in how to process or categorise the individual pieces of data. In supervised learning, dependent variables are predicted from independent variables by a mapping function. In contrast, unsupervised learning does not require a specific outcome variable, ML models will be explained in more details in the following Chapter 3.

ML has been shown to be an effective method for predicting several addictive behaviours in users, based on the processing of environmental data. Example behaviours include gambling, alcohol, cocaine and smoking (Moon et al., 2012; Mak et al., 2019). A systematic review by Mak et al. (2019) identified that supervised learning methods (e.g. Decision Tree (DT), Naïve Bayes, Logistic Regression, Support Vector Machines (SVM), Neural Networks) were used more often than unsupervised learning methods in these applications and shown a better prediction accuracy.

ML proved to offers the tools to extract smoking patterns from both internal and external craving factors (Abo-Tabik et al., 2019). There have been several ML algorithms that have been used to understand smokers' behaviour and the urge to smoke as it changes during the quitting period. Koslovsky et al. (2018b) demonstrated that feeding EMA self-reported data into a Bayesian Variable Selection for Multistate Markov Model methods could be effective for modelling the transition between different stages of the quitting processes in specific populations, however, their model was tested on data that was collected from 146 smokers at random intervals four times a day (i.e., not in real time) and retrospectively examined whether this could be used to predict relapse. While their formula produced good results, it is not clear whether it could be

used in real time situation to predict cravings that could lead to relapse.

Research by Zhang et al. (2019b) predicted daily smoking time. The research reported that XGBoost achieved 84.11% accuracy in predicting daily smoking time using 15,095 smokers' data from 2015 China Adult Tobacco Survey Report (there wasn't any use of any real behaviour data). The research modelled an equation that simulates the probability of smoking time based on smoker's demographic (e.g. gender, age, and average daily smoking). The researchers extracted and used statistical information from the dataset as well as some additional extracted features as input to the XGBoost model. The relying on a simulated dataset, makes the model lack in its understanding of a real daily smoking time, and the different external factors that my influence the smokers daily behaviour.

Few studies also made use of personalised intervention informed by supervised ML algorithms. For example, Williams et al. (2018) used targeted messages based on 256 participant's demographic profiles and previous self-reported smoking behaviour. Results showed that users who received these messages showed improved motivation to adopt a healthier lifestyle.

An interesting attempt in using supervised ML method is that of Dumortier et al. (2016), who used historical data collected from 349 university students to evaluate the urge to smoke at quit smoking period, based on 41 self-reported features (e.g., alcohol consumption, mood status, hunger, location, type of work, etc.) that might trigger an urge to smoke. Comparing three different ML algorithms (Naive Bayes Classifier, Discriminant Analysis Classifier, and DT), results revealed that ML had the ability to estimate the smokers' urge-rating with high accuracy in general, and the DT performed best with up to 86% accuracy. While this is encouraging, it is impossible to imagine smokers regularly reporting 41 features, and so more advanced combination of ML and variables extracted from the environment automatically should be considered.

While it is clear that ML combined with up-to-date technological advances such as those provided by mobile phones can provide a perfect platform for delivering targeted, personalised, and timely interventions, non of the available apps to date have made full use of all the techniques summarised in this chapter. However, limited attempts have been made ; for example, MapMySmoke (Schick et al., 2018) used a combination of EMA and ML to predict potential urges to smoke based on geo-location. The ML in the app was trained on pre-quit data collected from 4 participants, where they logged every smoking and craving events and was fed into a Hidden Markov Model to provide targeted intervention based on smokers' behaviour, including timing and places

in which individuals were most likely to smoke. The app then used these patterns and real-time information to provide timely and context-relevant support messages during the quit period. The study did not report any analytical results that are related to the Hidden Markov Models, but the feedback from participants was positive, suggesting that implementing smart apps to support users may not only be beneficial but also improve users' experience. However, the study also highlighted the risks associated with in-efficient predictions, as some participants reported receiving messages in the wrong timing, which had the undesirable effect of reminding them about their smoking habit hence awakening a craving, rather than helping them quit.

The research , described in the thesis follows the process of developing a deep learning model that could be used to develop a smart smoking cessation app. The idea of the model is to combine an EMA method with CTMoS, to predicts smoking behaviour based on both internal and external factors. The control theory model of smoking by (Bobashev et al., 2017) has been selected as the basic starting point, due to its ability to model nicotine level using the pharmacokinetic equation.

2.4 Conclusions

In conclusion, there are several approaches were used in the development of smoking cessation apps, mostly relay on using self-reporting of craving factors. A more advance approach that proved to be efficient is Auto-collected EMA smoking cessation app, the new method minimized the reliance on self-reported data and only collect self-reported smoking events at pre-quitting period. This approach is better than Regular EMA smoking cessation app but it needs advanced data analysis when collecting more than one craving factor. There are few attempts in using ML in modelling smoking behaviour and automatic prediction of smoking events, mostly used data that can be only self-reported by the smoker and this did not solve self-reporting problem. Only one attempt reported the use of Auto-collected smartphone data with ML model but the study did not report any analytical results about the performance of the ML model.

Chapter 3

Introduction to machine learning

3.1 Introduction to machine learning and reproductive data analysis

Machine Learning (ML) is a field that is based on the principle of learning by feature analysis and pattern extraction using a set of pre-collected data in order to master a particular task. ML models have been used widely in many fields due to their ability to perform complex tasks with minimum effort (Ray, 2019). These problems can be either classification, in which the model task is assigning the input sample to a category; or the regression problem where the input samples will produce continues sample data (Chen et al., 2017).

The design of ML models can be either supervised or unsupervised. In supervised learning models, the target is known, such that the targeted outputs from the training set are pre-determined and can be predicted from independent variables by a mapping function. For example, the use of supervised learning methods in price prediction, in which the historical data is used to predict an increase or decrease in price. In contrast, in unsupervised learning the learning is done on an unlabelled dataset; meaning a specific outcome is unknown, but dimensionality reduction and/or clustering processes are applied to the independent variables. For example, some websites use unsupervised learning to group users for a group-specific recommendations system (Sharma and Nandal, 2019).

All ML models have different parameters and hyperparameters, and these are varied in number and the type between different model types and architectures. In general

hyperparameters are pre-set during the design stage, while the parameters are optimised during the learning (training) process.

After the training phase, the model needs to be evaluated for its reliability in doing the target task; this is called the testing process. A model is evaluated based on its prediction accuracy. A model accuracy is an evaluation measure to decide the model's ability in identifying relationships and patterns between input features in a dataset, in which it is measured as the percentage of correctly predicted samples to the overall samples. In order to calculate a model accuracy for a binary classifier, four important measures should be computed. These are,

1. True Positive (TP): the number of rightly predicted values from the positive class (i.e. the event that the model tries to predict and usually coded as 1).

2. True Negative (TN): the number of rightly predicted values from the negative class (i.e. which is the second class and usually labelled as 0).

3. False Positive (FP): the number of positive labels that are wrongly predicted.

4. False Negative (FN). the number of negative labels that are wrongly predicted.

Accuracy is then calculated using the following equation (Ikram and Cherukuri, 2016; Thaseen and Kumar, 2017),

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Precision is way to measure the model performance in protecting positive class, which is known as the ratio between the TP and the total number of samples classified as positive class. Recall is used to measure the model accuracy in detecting positive samples, and it is calculated as the ratio between TP and total number of positive samples (Chakravarthi et al., 2020),

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

Another popular model for evaluation measure is f1-score. f1-score is the harmonic mean of precision and recall. f1-score is an important evaluation matrix for models with imbalanced labels (i.e. when the number of samples for one class is significantly larger than the number of samples of the second class). f1-score is usually used when

the FN and FP are crucial in evaluating the model, and is suitable for use in the case of smoking event prediction, where due to the unbalanced nature of the data (i.e., even among heavy smokers, many more minutes of the days are labelled as 'non-smoking' compared to 'smoking'). (Jeni et al., 2013; Wardhani et al., 2019).

$$Precision = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.4}$$

A common practice is to test the model on an unseen portion of the dataset during the training process (Hutter et al., 2019). This means, for model evaluation, the first step is to partition the dataset into two portions, namely training and testing sets. The training set should be further separated into training and validation set, this is usually done to avoid data leakage problem. A good model is one that shows high accuracy in both training and testing processes and which is robust to variation in hyper-parameters. However, sometimes even if the model performs well in the training process, it shows degradation in its performance when it is tested; this situation is called over-fitting. In most cases, over-fitting occurs when the model is over-trained; and sometimes if the model is too complex, it also will suffer from over-fitting. In over-fitting, the model becomes over specialised to the training set, and as a result it loses its generality and its ability to capture the relations within previously unseen data (Erickson et al., 2017).

The most common way for model evaluation is called the holdout method, Figure 3.1(a); here, the data will be portioned to training, testing and validation. The percentage of each portion of the dataset will be set either randomly, or sometimes by testing to see what minimum size is suitable for good training without affecting the accuracy of the testing. This method is inefficient for small datasets, where the model, in most cases, will be biased depending on the splitting ratios.

To overcome this problem, the k-fold method is used, Figure 3.1(b). In this approach, the holdout method is repeated k times. This means the dataset is first portioned into k portions. At each round, one of the portions will be used for testing, and the remaining k-1 portions will be used for training and validation. This means that each sample in this method will be trained k-2 times and tested/validated at least once. This approach significantly reduces any bias since all the data at some point will be used for training. Furthermore, this method reduces the variance because all the data is used in the test/validation process (Xiong et al., 2020). The only concern in this method is the time, as it takes too long to test all data and that makes it inefficient when the dataset is very large.

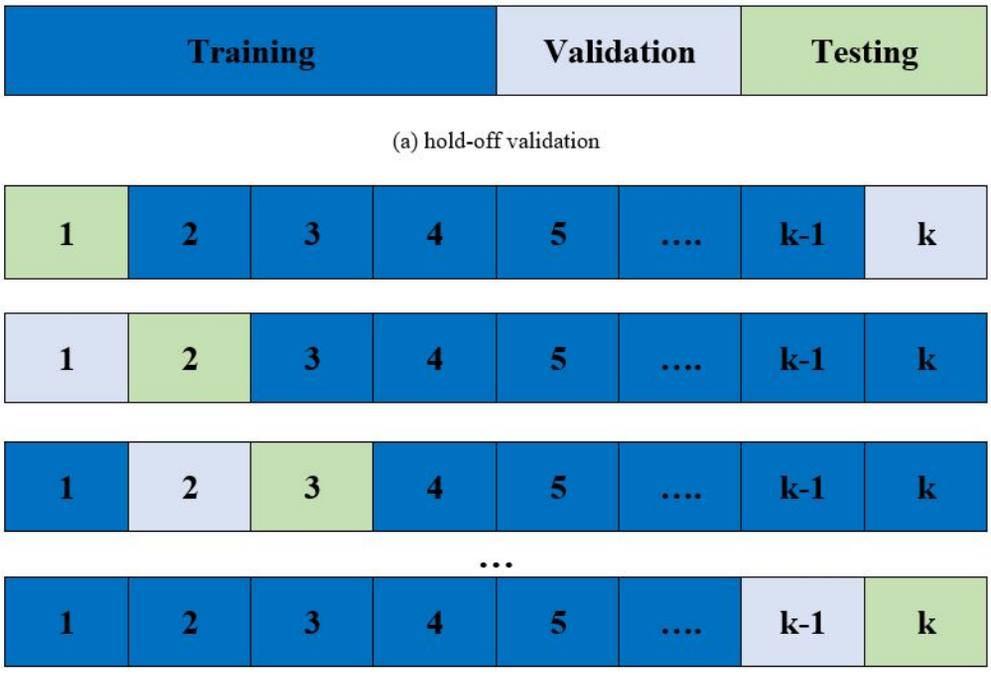


Figure 3.1: The most common validation approaches. The dark blue is the training set, the light blue is the validation, and the green is the test set based on Xiong et al. (2020).

3.2 Classical ML models

The next section will highlight and briefly describe two commonly used ML methods. These methods are well known for their efficiency in solving different machine learning problems. These algorithms will be used later for the experiments in this thesis.

3.2.1 Support Vector Machine (SVM)

SVM is a supervised ML method proposed by Vapnik (2013). SVM uses optimisation theory along with statistical learning to solve classification and regression problems. It is well known for its efficiency and ability to avoid overfitting (Chen et al., 2017; Kalantar et al., 2018), and it can accomplish better performance than other ML algorithms using small dataset or high dimensionality (Tao et al., 2018). However, despite all these advantages, SVM suffers from the problem of a long training time, and high error rate when target labels overlap (Ikram and Cherukuri, 2016; Thaseen and Kumar,

2017).

The SVM classifier is based on the principle of finding a hyperplane that can separate the samples from different classes (Meenal and Selvakumar, 2018), as shown in Figure 3.2. It uses the functional margin (i.e. the separation space between the training points and hyperplane) to assign samples to each class. The highest possible functional margin gives the best data point separation, the points with the lowest distance to the hyperplane are called support vectors. For feature vector x which has i features and y as output binary class, the SVM decision function is calculated

$$\begin{aligned} f(x) &= \left(\sum_{i=1}^{n} \alpha_i y_i x_i\right)^T x + b \\ &= \sum_{i=1}^{n} \alpha_i y_i \langle \acute{x}_i, x \rangle + b \end{aligned} \tag{3.5}$$

where α is the Lagrange multiplier, $\langle , \rangle$ is the inner product operation, and b is the bias value.

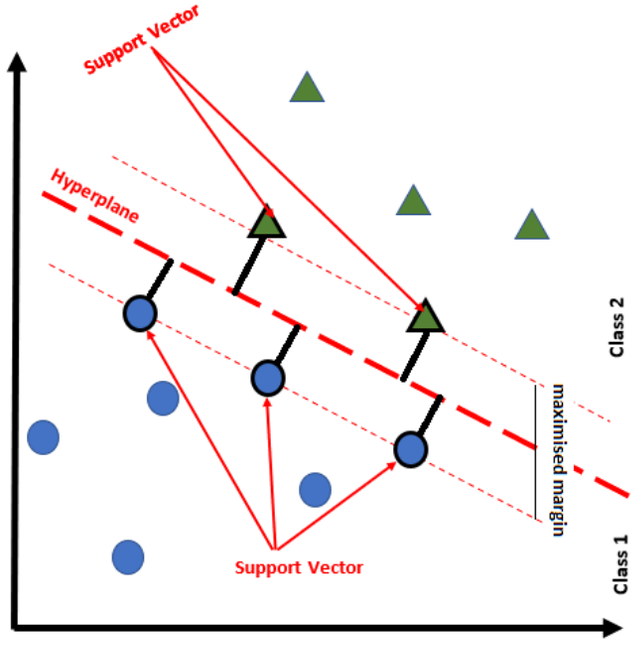


Figure 3.2: An example showing a linear SVM classifier (Wang et al., 2018; Zidi et al., 2017).

However, non-linear discrimination data will need a more complex separation method

to map the data points to a higher-dimensionality space, see Figure 3.3. Thus, the detection function

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i \langle \emptyset(\acute{x_i}), \emptyset(x) \rangle + b \tag{3.6}$$

is used, where $\emptyset$ is the mapping function from low dimensional to higher dimensional plane.

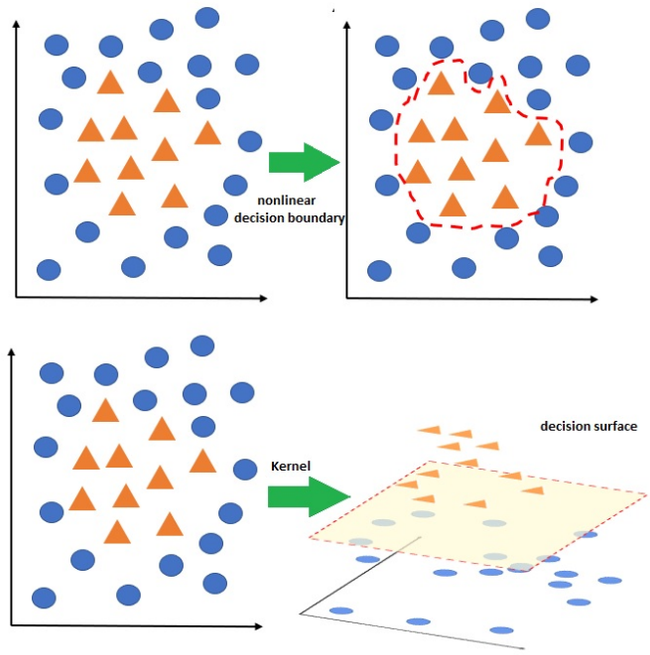


Figure 3.3: An example showing a non-linear SVM classifier (Wang et al., 2018; Zidi et al., 2017).

Now, in the higher-dimensional space, the hyperplane will be defined as a set of points with constant dot product in the presences of the vector. This means that the kernel function will take the input vector and returns the dot product of that vector in the feature space,

$$K(x, z) = \langle \emptyset(x), \emptyset(z) \rangle. \tag{3.7}$$

The low dimensional kernel $K(x_i, x)$ can be easily used to calculate the inner product in the higher dimensional kernel. The kernel function is usually calculated as $\langle \emptyset(x), v(x) \rangle$ (Wang et al., 2018). In this sense, the functional margin and kernel parameters will both determine the accuracy of the classifier (Meng et al., 2019; Wang et al., 2018).

3.2.2 Decision Tree Method (DT)

A DT is a ML model that can be effectively employed to solve both classification and regression problems. It uses a top-down repetitive partitioning approach to generate the tree structure (Barros et al., 2011). DT uses information theory techniques to analyse a large amount of data for automated data portioning. The root is the most important part in the tree; it should be the most informative node in the whole tree. The root-node is then followed by several nodes (Long and Wu, 2012). DT models may use different measures such as gain ratio (Quinlan, 1993), information gain (Quinlan, 1986), distance-based measures (De Mántaras, 1991), and Gini index (Breiman et al., 1984) for data separation.

Classification and regression tree (CART) is a popular form of DT that can be employed for both classification of regression problems (Breiman et al., 2017); it used the Least Squared Deviation (LSD) as its impurity function (i.e a function that measures the homogeneity of the labels at the child node) (Vanli et al., 2019),

$$R(t) = \frac{1}{N_w(t)} \sum_{i \in t} w_i f_i \left(y_i - \bar{y}_i(t)\right)^2 \tag{3.8}$$

$$\bar{y}(t) = \frac{1}{N_w(t)} \sum_{i \in t} w_i f_i y_i \tag{3.9}$$

$$N_w(t) = \sum_{i \in t} w_i f_i \tag{3.10}$$

where y_i is the response and $\bar{y}_i$ is the value of the mean response, w_i is the calculated weight value for each i and f_i is the recorded response. $N_w(t)$ is the weighted number of samples in node t. The splitting criterion is calculated using

$$Q(s,t) = R(t) - R(t_L) - R(t_R) \tag{3.11}$$

in which t_R is the right child and t_R is the left child of the node t. The split criteria is determined to maximise $Q(s,t)$.

=When a set of DTs are combined, it forms an ensemble model; this approach is typically used to group a set of weak models to form one strong model, examples of DT ensemble are bagging (Breiman, 1996) and boosting (Freund and Schapire, 1997). In Tree Bagging, the data set is partitioned into several subsets; each subset is used to train its own tree, and the final model is a combination of all these trained sub-models. A plurality vote between the sub-trees is used for predicting the class (Tu et al., 2009; Zhu

et al., 2012). Tree Boosting, on the other hand, uses an iterative re-training method. It is based on an equational training process, meaning that at each step the newly trained model learns from the error of the previous classifier by increasing the weight as the training progress (Bonissone et al., 2010). While Boosting improves the accuracy of the prediction of the classical DT, it is also very slow in learning, sensitive to noise, and it increases the possibility of over-fitting (Briem et al., 2002).

Random forest (RF) is another form of the ensemble DT; this model is efficient because it reduces the over-fitting problem. This method is a modified version of Bagging, but instead of using all subsets to generate multiple versions of the DL, some of the samples are left out (out-bagging) then used in cross-validation to estimate classification accuracy while avoiding over-fitting (Mim and Zamil, 2018). The error generated during the validation process is called out-of-bag (OOB) error. The output is the averaging probability of all predicted class generated by all the trees (Belgiu and Drăguţ, 2016).

3.3 Artificial Neural Networks and Deep Learning

Artificial Neural Networks (ANN) are a subfield of ML; they are a form of non-linear mathematical model that simulates the behaviour of biological neurons. The operation of each neuron is defined by Equation 3.6 as with the SVM. However, an ANN will have multiple neurons at multiple levels all with the same activation function, which gives a non-linear transform of the output of the neuron. In contrast, an SVM will have a single node, with a non-linear transform of the input-data to similarly code complex decision boundaries. The early architecture of ANNs was a fully connected network of neuron, otherwise known as the multi-layer perceptron (MLP). Although the MLP outperformed other ML methods in both classification and regression problems in many applications, it failed to accomplish a good result when it was used on data with a complex feature set (Huang and Kuo, 2018).

A more advanced form of ANN is a deep neural network (DNN); an ANN with multiple hidden layers, where each layer processes the output of the previous layer. This multi-layered architecture of Deep Learning (DL) models has been used to resolve many complex problems such as image and speech recognition, and achieved better results compared to previous ANN and other ML approaches. DL increases the capability to capture complex data patterns, and exploit non-linear information from a large amount of data (Bouktif et al., 2018; Ordóñez and Roggen, 2016; Popa et al.,

2019; Zahid et al., 2019).

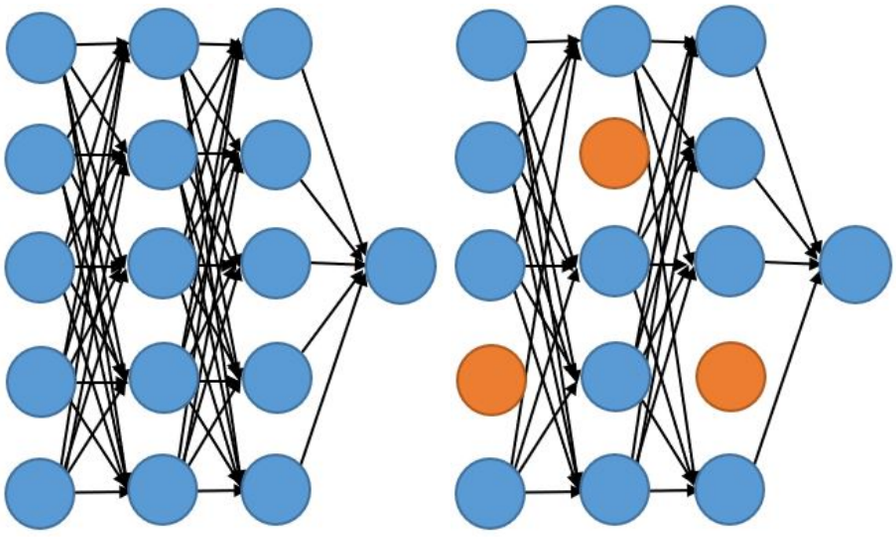
This thesis will explore the possibility of building a prediction model of smoking behaviour using DL architecture. The aim of using this advanced technique is to capture the complex patterns of influences on smokers' behaviour using data collected from sensors on their smartphone devices.

3.4 Techniques to improve the DL model

Because of the large number of layers, and complex design of DL model, several problems may appear during the migration of processed information along the network's hidden layers. Many techniques have been used to improve the performance of the network, and avoid losing important information while learning. Some are outlined in Sections 3.4.1 and 3.4.2.

3.4.1 Drop out

This method was first proposed by Hinton et al. (2012), and it applies random masking noise to enlarge the dataset, which creates random variation in the input data to avoid over-fitting (Zhao et al., 2017). The basic idea of drop out is that the network will randomly discard some of the network units based on a percentage assigned at the network design stage. It is important to mention that drop out does not apply at the validation process, and is turned off during testing, meaning all the data will be processed by all neurons (Cao et al., 2019). See Figure 3.4 for an example of the process.



(a) The network before applying dropout.

(b) The network after applying dropout.

Figure 3.4: Applying the dropout method on DNN.

3.4.2 Batch Normalisation

The multilayered network of DL model makes training the network a very challenging process. It is sometimes attributed to the change in the distribution of each batch of inputs along the layers when the weights are updated during the learning process; this problem is called internal covariation. Internal covariation usually leads to slowing of the training process by demanding a low learning rate and careful parameter initialisation (Ioffe and Szegedy, 2015).

To solve this problem, there is a need to reduce the internal covariation shift by fixing the inner layers' batch inputs during the training process. Batch normalisation offers a solution for increasing the DL stabilisation, by normalising the output produced from one layer before it passes to the next layer. To this end, it first subtracts the batch mean, followed by dividing by the batch standard deviation. For a layer with inputs $x = \left\{ x_1, \quad x_2, \quad \dots \quad x_n \right\}$; batch normalisation can be calculated,

$$\hat{x}_i = \frac{x_i - E\left[x_i\right]}{\sqrt{\text{Var}\left[x_i\right]}} \tag{3.12}$$

where $E\left[x_i\right]$ and $\text{Var}\left[x_i\right]$ are the mean and variance respectively for each i unit. To avoid the change in layer representation that occurs due to the normalising of each

input value in a layer. Two parameters γ and β are used by all neurons as the learning configuration parameters to shift and scale the normalised value,

$$y_i \ = \ \gamma * \hat{x}_i + \beta \tag{3.13}$$

where $y = \Big\{ y_1, \quad y_2, \quad \dots \quad y_n \Big\}$ is the output, where γ and β are automatically learned from the DL network, and restore the original distribution learned from the previous layer. This will ensure the stability of the data distribution during the training process (Wu et al., 2019).

3.5 1D Convolution Neural Network

Convolutional neural networks (CNN) are a well-known neural network architecture that has been used since the 1980s (Kalchbrenner et al., 2014; LeCun et al., 1998). CNN has been proved its efficiency in many applications (Nguyen and Grishman, 2015). This is due to its high ability to learn and extract feature vectors from non-linear and complex datasets, and also its ability to be applied in parallel computational architectures (Sorokin et al., 2018). This reflects its use of multiple, independent calculations.

1D-CNN is one form of CNN that performs a convolutional operation on the raw input data. The CNN network is designed in a way that it learns using a hierarchical feature extraction process. The CNN is usually made of a sequence of convolutional layers, each of which uses a set of sliding windows (kernel) and filters. The number of layers, size of the kernels and the number of filters are all presets during the network design stage.

The convolutional operation slides (shifts) the kernel across the data-vector, stopping at each point, where a local matrix multiplication is applied and the result is summed into the final feature map. This in turn reduces the need for complex feature extraction procedures, often used by other ML models, and ensures better results by enabling the model to extract its feature map (Fu et al., 2019). The convolution process for m layers is

$$y \ = \ F(X|\Theta) \ = f_L(...f_2(f_1(X|\Theta_1)|\Theta_2)|\Theta_m) \tag{3.14}$$

where m is the number of hidden layers, y is the predicted output, X is the set of inputs,

and Θ is the layer related set of parameters. The convolution operation for layer i is

$$y_i = f_i(X_i|\Theta_i) \; = \; h(W \otimes X_i + b) \; \Theta_i = [W, b] \, , \tag{3.15}$$

where $\otimes$ is the convolution operation, and W and b are weights and bias respectively (Abdoli et al., 2019; Deka et al., 2019).

One problem with the feature map produced by convolution layer is that it heavily relies on the location of the significant features in the input. This means that any small transition in a feature position in the input data between samples causes different feature map (Kim and Cho, 2018, 2019). A common practice is to add a pooling layer after each CNN layer. The pooling layer is based on down-sampling approach, where it creates an a new version of the input sample that contains all the important information, without the unnecessary data that may not be as useful to the task. This means that it reduces the dimensionality of the convolution layer feature map, which improves the feature extraction process and reduces unnecessary calculations (Huang et al., 2020; Phan et al., 2016).

In most CNN networks, following the CNN+pooling layers, a final classification layer is usually used. This can be any ML classifier such as SVM, but the most well-known approach is to use a fully connected ANN as a final layer in the network. Using fully connected ANN can simplify the training process (Niu and Suen, 2012; Xue et al., 2016). Figure 3.5 shows the general 1D-CNN design.

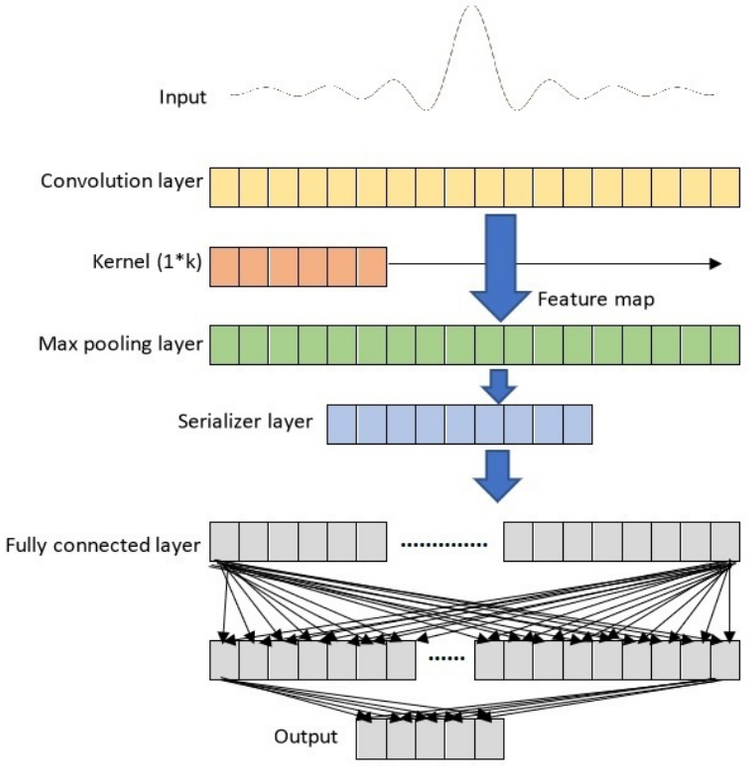


Figure 3.5: General 1D-CNN design based on Kim et al. (2020)

3.6 Recurrent Neural Network and Long Short-Term Memory

A Recurrent Neural Network (RNN) (Rumelhart et al., 1986) is a modified version of feed-forward ANN, in which the output at each point is linked to all previous inputs. This means that each layer in the current hidden state is a function of the current input and the previous hidden state. For the input sequence $x = [x_1, \ x_2, \ x_3, \ \ \ldots x_n]$ using RNN $h_t = f(h_{t-1}, x_t)$, where x_t is the input at time-step t , and h_{t-1} is the previous hidden state (Bouktif et al., 2018). Figure 3.6 shows the general architecture of an RNN.

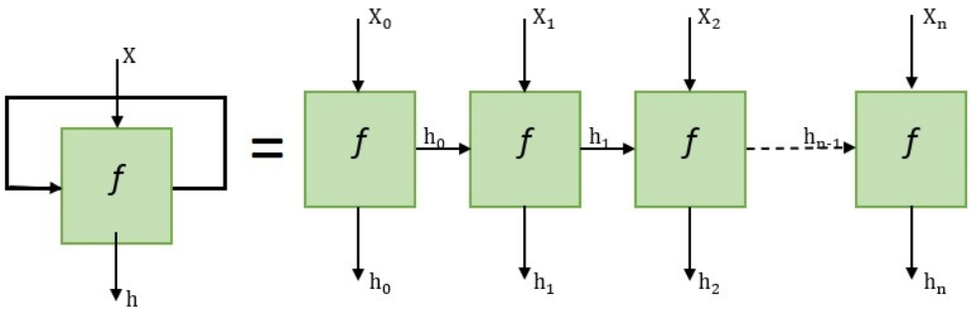


Figure 3.6: Basic architecture of RNN based on Zhao et al. (2017)

Although this architecture is useful in extracting patterns from time-dependent data samples, it can only look a few steps back, because of the vanishing and exploding gradient problem, which makes RNNs challenging to train sometimes (Zhou et al., 2016). The Long Short Term Memory (LSTM) is an improved version of RNN, as it has been designed in a way that can overcome long term dependencies problems. This network is trained using back propagation method (Yan et al., 2018).

Each LSTM layer is a set of blocks, where each block is consisting of several multiplicative units and memory cells that are recurrently connected. These memory cells are considered the main contribution in the LSTM architecture. The memory cell has three gates, and stores the information that is obtained at this step. It then either keeps it, release it, or reset it, based on the state of the controlling gate. The memory unit gates are called input, output, and forget gate; each gate is controlled by a sigmoid activation function $\sigma \in [0, 1]$ (Tian et al., 2018). Figure 3.7 shows the main structure of the LSTM memory unit.

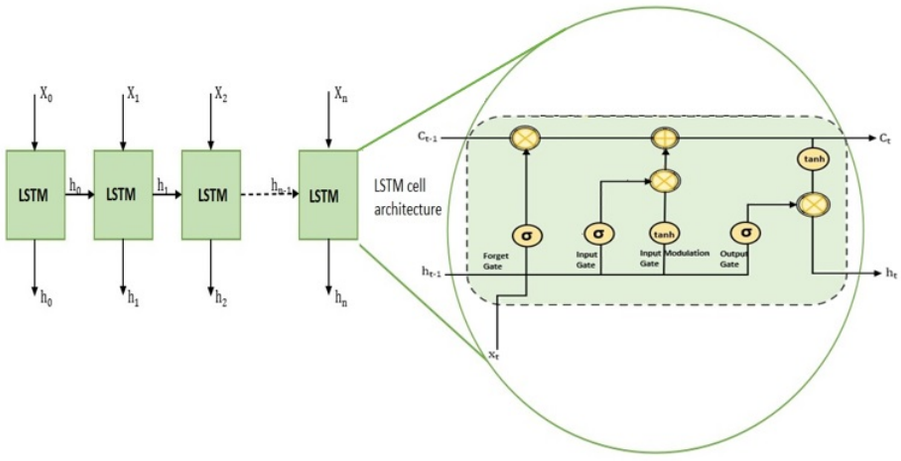


Figure 3.7: LSTM cell architecture based on Zhao et al. (2017)

Like the RNN, each gate receives an input X for the time t and the previous hidden state h_{t-1}. The forget gate f_t determines how much information will be kept from the previous state c_{t-1}; f_t is calculated using (Wang et al., 2019; Zhou et al., 2015),

$$f_t = \sigma\left(W_f \bullet [h_{t-1},\ X_t] + b_f\right) \tag{3.16}$$

where W_f, b_f are the weight and the bias for the forget gate. The input gate i_t, on the other hand, is responsible for controlling the amount of current information to be considered as input for generating the current state c_t ; i_t is calculated using the following equation; W_i, b_i are the weight and the bias for the input gate (Wang et al., 2019; Zhou et al., 2015),

$$i_t = \sigma\left(W_i \bullet [h_{t-1},\ X_t] + b_i\right). \tag{3.17}$$

Now the current hidden state c_t will be calculated using the long-term information obtained from f_t and the short-term information from i_t, as in the following equations, W_c, b_c are the weight and the bias for the current state (Wang et al., 2019; Zhou et al., 2015),

$$\widetilde{c}_t = tanh(W_c \bullet [h_{t-1},\ X_t] + b_c) \tag{3.18}$$

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c}_t \tag{3.19}$$

where $\tanh(\bullet)$ is the activation function, and $*$ is an element-wise product. The last gate o_t is the output gate, which will decide the amount of information to be treated as

output, and will be calculated using

$$o_t = \sigma\left(W_o \bullet [h_{t-1},\ X_t] + b_o\right) \tag{3.20}$$

where W_o, b_o are the weight and the bias for the output gate. Finally, since all gates are controlling the information flow using the element-wise product; the final out h_t will be calculated (Wang et al., 2019; Zhou et al., 2015) as,

$$h_t = o_t * tanh(c_t). \tag{3.21}$$

A modified version of LSTM is the Bidirectional LSTM (Bi-LSTM) (Schuster and Paliwal, 1997). Bi-LSTM consists of two forward and backward LSTM models, with both models connected to the same output (Rao et al., 2018), see Figure 3.8. In the Bi-LSTM two copies of the input pass to the model, where it gives one copy to the forward LSTM, and one to the backward LSTM.

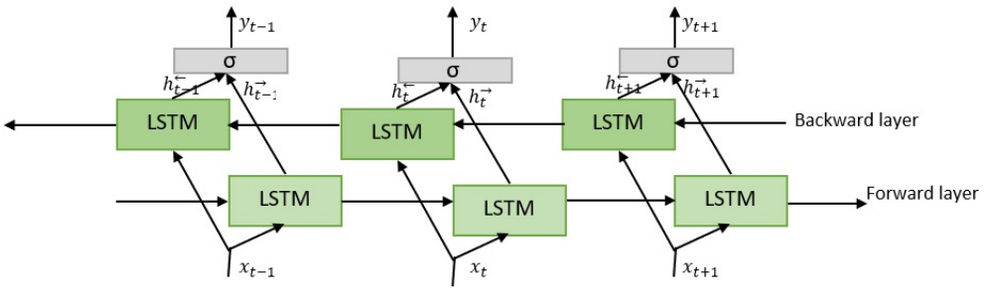


Figure 3.8: Illustration of BiLSTM network based on Zhao et al. (2017)

In the BiLSTM f_i,i_t, and o_t, are all calculated for both the forward and backward networks; $h_t^{\rightarrow}$ for the forward LSTM network will be calculated using the current t and the previous $t-1$ using (Zhao et al., 2017),

$$f_t^{\rightarrow} = \sigma\left(W_f^{\rightarrow} \bullet \left[h_{t-1}^{\rightarrow},\ X_t^{\rightarrow}\right] + b_t^{\rightarrow}\right) \tag{3.22}$$

$$i_t^{\rightarrow} = \sigma\left(W_i^{\rightarrow} \bullet \left[h_{t-1}^{\rightarrow},\ X_t^{\rightarrow}\right] + b_i^{\rightarrow}\right) \tag{3.23}$$

$$\widetilde{c_t}^{\rightarrow} = tanh\left(W_c^{\rightarrow} \bullet \left[h_{t-1}^{\rightarrow},\ X_t^{\rightarrow}\right] + b_c^{\rightarrow}\right) \tag{3.24}$$

$$c_t^{\rightarrow} = f_t^{\rightarrow} * \widetilde{c}_{t-1}^{\rightarrow} + i_t^{\rightarrow} * \widetilde{c_t}^{\rightarrow} \tag{3.25}$$

$$o_t^{\rightarrow} = \sigma\left(W_o^{\rightarrow} \bullet \left[h_{t-1}^{\rightarrow},\ X_t^{\rightarrow}\right] + b_o^{\rightarrow}\right) \tag{3.26}$$

$$h_t^{\rightarrow} = o_t^{\rightarrow} * tanh(c_t^{\rightarrow}). \tag{3.27}$$

Meanwhile the $h_t^{\leftarrow}$ for the backward LSTM network will be calculated using the current t and the next $t+1$ as

$$f_t^{\leftarrow} = \sigma\left(W_f^{\leftarrow} \bullet \left[h_{t+1}^{\leftarrow},\ X_t^{\leftarrow}\right] + b_t^{\leftarrow}\right) \tag{3.28}$$

$$i_t^{\leftarrow} = \sigma\left(W_i^{\leftarrow} \bullet \left[h_{t+1}^{\leftarrow},\ X_t^{\leftarrow}\right] + b_i^{\leftarrow}\right) \tag{3.29}$$

$$\widetilde{c}_t^{\leftarrow} = tanh\left(W_c^{\leftarrow} \bullet \left[h_{t+1}^{\leftarrow},\ X_t^{\leftarrow}\right] + b_c^{\leftarrow}\right) \tag{3.30}$$

$$c_t^{\leftarrow} = f_t^{\leftarrow} * \widetilde{c}_{t+1}^{\leftarrow} + i_t^{\leftarrow} * \widetilde{c}_t^{\leftarrow} \tag{3.31}$$

$$o_t^{\leftarrow} = \sigma\left(W_o^{\leftarrow} \bullet \left[h_{t+1}^{\leftarrow},\ X_t^{\leftarrow}\right] + b_o^{\leftarrow}\right) \tag{3.32}$$

$$h_t^{\leftarrow} = o_t^{\leftarrow} * tanh(c_t^{\leftarrow}) \tag{3.33}$$

The final h_t will be the concatenated vector of both the $h_t^{\rightarrow}$ and $h_t^{\leftarrow}$,

$$h_t = h_t^{\rightarrow} \bigoplus h_t^{\leftarrow}. \tag{3.34}$$

The use of Bi-LSTM may not be useful in all time-sensitive models. However, it gives good results when the problem needs a full understanding of the previous and the following information for every point in the data sequence (Tavakoli, 2019).

3.7 DL models activation function

An activation function is an important part of the DNN architecture. This function is responsible for determining the network output and hence, its accuracy.

Softmax (Huang et al., 2018; Tang, 2013) is a widely used activation function in DL classification models. The softmax function assigns a discrete probability distribution for K classes that can be calculated by,

$$P_K = \frac{exp(o_k)}{\sum_{K=0}^{n-1} exp(o_K)} \tag{3.35}$$

$$o_K = \sum_{i}^{n-1} W_i X_i \tag{3.36}$$

Where W and X are the weights and the input of the hidden layer, the predicted class y is calculated by

$$y = \underset{i \in 1,\ldots N}{arg\ max}\ P_i \tag{3.37}$$

Another well-known activation function is ReLU (Rectified Linear Unit) (Agarap, 2018; Zhang et al., 2019a; Zou et al., 2020), that proved its efficiency over traditional neural network (logistic sigmoid, hyperbolic tangent) activation functions (Camps et al., 2018). ReLU is a ramp function and it is similar to half-wave rectification in engineering, it is presented by $f(x) = max(x, 0)$, where the out is a linear function when $x \geq 0$, or 0 when the output $x < 0$.

The select of the activation function highly depends on the target design of the DL network.

3.8 Using genetic algorithms to select the right DL design parameters

Any DL model architecture has a set of hyperparameters, and these vary between different model types. For example, the number of units in the LSTM and fully connected ANN, or filter size in CNN model, etc. Finding the optimum value for each one of these parameters is a challenging problem, as it requires a long time to search an unlimited number of combinations (Hutter et al., 2019). The Genetic Algorithm (GA) was first introduced in a paper by (Fraser, 1957), then later it was first developed by Holland (1975). It is based on the concept of natural selection and natural genetics. GA uses the concept of "Survival of the fittest" to obtain the optimal solution (Meng et al., 2019). In a GA a set called the population, represents the set of solutions. Each member in the population is called a chromosome. The mutation and crossover operators are applied to the population set to maintain population diversity and prevent local optimum (see Figure 3.9).

In the crossover operation, part of the parent chromosomes is replaced and combined to generate a new member. In contrast, mutation is the process of changing individual chromosomal genes in a parent chromosome, to generate a new individual offspring. The individuals in the population are chosen based on a selected fitness function, such that the individual with the higher quality will have a greater chance to be selected to generate the next generation (Tao et al., 2018).

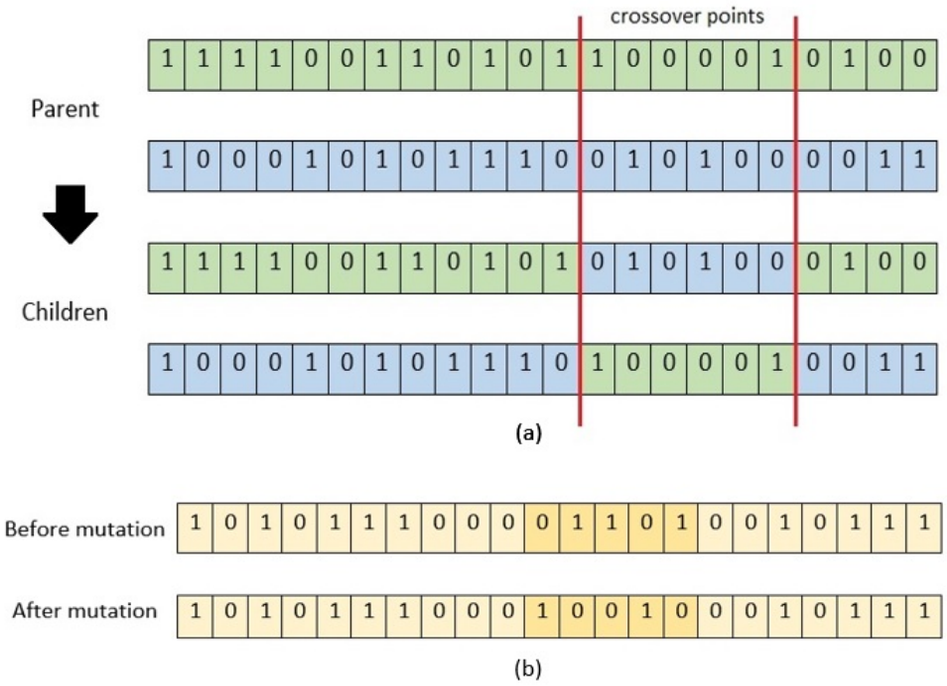


Figure 3.9: Example of the effect of applying crossover and mutation on population members (a) crossover operator and (b) is mutation operation.

GA can provided a good solution to improve the search process for a better DL model, and it has been employed before for hyperparameter optimisations for ML models (Meng et al., 2019; Sukawattanavijit et al., 2017; Tao et al., 2018; Tao et al., 2019). It has also been employed to select the best values for model architecture in a DL problems (Bouktif et al., 2018, 2020; Lu et al., 2020; Sun et al., 2020).

3.9 Conclusions

This chapter has explained in detail different machine learning methods that will be used later in this project. It started with a general explanation of what is ML and what validation and evaluation approaches are commonly used. Then it went into detail on two common ML methods, SVM and CT.

The chapter also explained relevant DL methods and some techniques to improve the DL models. It also explained that both 1D-CNN and LSTM methods, which will be used later to develop the smoking behaviour model. At last, the chapter gave a brief description of GA, which will be used for hyperparameters selection.

Chapter 4

Datasets description and analysis

This research designs a deep learning model that can predict aspects representing the internal drivers of smokers' behaviour (e.g. nicotine level, craving). The model is based on predicting smoking events using external factors (e.g. location), which can be collected using the smoker's personal mobile device.

In order to accomplish this task, two data sets will be used in this project. Both datasets used EMA methods to collect real-time data samples from participants using their personal mobile devices. Firstly, the QSense dataset (Naughton et al., 2016) with the aim of using this dataset for a second validation of the Control Theory Model of Smoking (CTMoS) on human smokers' data; CTMoS was first primarily validated on animals data (Bobashev et al., 2017). The QSense dataset contains a self-reported craving level associated with real-time self-reported smoking events and other emotional and environmental factors, making it perfect for testing the reliability of simulated internal nicotine craving in relation to cue-induced smoking craving.

The second dataset is a Smartphone Sensors dataset, collected during the course of this PhD project. This dataset contains passively collect environmental data (location and motion data) along with with real-time self-reported smoking events. The aim of collecting this data was to enable the evaluation of a deep learning model that automatically predicts smoking events in real time, by combining a mathematical model of internal factors with external environmental factors, both of which are known to influence smoking behaviour. This dataset is unique in its characteristic (automatically collected data in real time) and its extensiveness (1-sample per minute for 14 days).

This chapter will first provide an explanation of the QSense dataset. Then use this dataset to validate the CTMoS, after a detailed description of how CTMoS can model internal craving factors. Second, the chapter will also give a deep explication of the

Smartphone Sensors dataset including study design and type of data collected, etc.

All novel data used in this project was collected following standard ethical applications protocols, and approved by the Manchester Metropolitan University (MMU) Research Ethics and Governance Committee.

4.1 QSense dataset study design

The QSense dataset (Naughton et al., 2016) was not collected during this PhD, and was initially approved by the Cambridge Psychology Research Ethics Committee Application No.: Pre.2014.65, and NHS Health Research Authority IRAS project ID: 159828. The QSense project collected data from current smokers, aged 16-70 years old, who were interested in quitting smoking, and were willing to set a quit date (so that data could be collected before and after quitting). The Qsense data was collected using an app designed for Android mobile phones.

Following informed consent, participants were asked to install and use the QSense app for 14 days before quitting smoking (pre-quitting period). During this period, participants were asked to report each cigarette they smoke in real-time, along with filling a short assessment that included: smoking craving, stress, depress levels, the presence of others people if they are around other smokers, and their current location (home, work, socializing, or others). This data was collected alongside an automatically detected GPS location, indicating where participants logged the smoking event. Table 4.1 provides details of the QSense data.

Table 4.1: Explanation of the QSense data that are used in this project

Collected data	**Description**
User id	A unique number that identifies each participant.
Craving rating	1=”No urges”, 2=”Slight”, 3=”Moderate”, 4=”Strong”, 5=”Very Strong”, 6=”Extremely Strong”.
Stress rating	1= ”Not at all”, 2=”Slightly”, 3=”Somewhat”, 4=”Very”, 5=”Extremely”.
Depressed rating	1= ”Not at all”, 2=”Slightly”, 3=”Somewhat”, 4=”Very”, 5=”Extremely”.
Current situation	1=”Home”, 2=”Working”, 3=”Socialising”, 4=”Other”
Others smoking	1= “I am alone”, 2=” Who is smoking”, 3= “Who are not smoking.”.
Who with	1= “Nobody”, 2=”Friends/Family”, 3=”Colleagues”, 4=”Others”.
Log time	Date and time of the smoking event.

The 14-day pre-quitting period was followed by post-quitting intervention period of 28 days, which started after the quitting date. During the intervention period, the app began to send messages to the user whenever they entered an area that had previously been registered as high smoking location risk. The app detected the participant location by collecting the GPS data every 15 minutes. While the current project used the pre-quit data, the 28 days post-quit data has not been used in this project.

4.2 Control theory of smoking

Bobashev et al. (2017) proposed a mathematical model which is based on opponent process principle (see section 2.1), it has been presented as a cascading feedback loop, as shown in Figure 4.1.

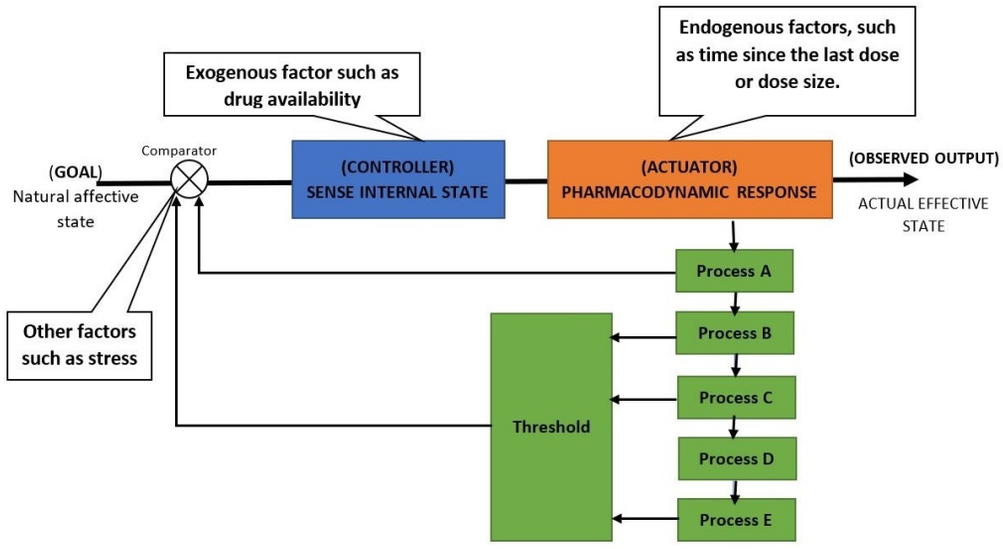


Figure 4.1: Control theory model of smoking based on Bobashev et al. (2017)

CTMoS does not model any neurobiological process, as it was developed with phenomenological interpretation in mind, with the following equations representing the cascading model functions;

$$ProcessA: \frac{dY_1}{dt} = e^{-\alpha t} - b_1 Y_1 \tag{4.1}$$

$$ProcessB: \frac{dY_2}{dt} = a_1 Y_1 - b_2 Y_2 \tag{4.2}$$

$$ProcessC: \frac{dY_3}{dt} = a_2 Y_2 - b_3 Y_3 \tag{4.3}$$

$$ProcessD: \frac{dY_4}{dt} = a_3 Y_3 - b_4 Y_4 \tag{4.4}$$

$$ProcessE: \frac{dY_5}{dt} = a_4 Y_4 - b_5 Y_5 \tag{4.5}$$

Each of these equations presents a weighted integration of the previous equation, causing the next process to be longer than the previous one. a, b, and α values are scaling coefficients, and all a, b, and α initial values are equal to zero.

Y_1 represents the effect of nicotine level, modelled with a pharmacokinetic equation. Y_2 represents the toxicity level and how the body processes the drug. Y_3 represents a daily smoking habit. Y_5 presents a longer scaling habit, which is scaled in years (rather than minutes/ hours/ days). While the process Y_4 has not been interpreted; it has been used to add a scaling period between $Y_3$and Y_5 which results in a slow change in process Y_5

To simulate the smoker's behaviour, a threshold value is defined to prompt self-administration (of nicotine, i.e. smoking). The threshold (T) is modelled as,

$$T = (\beta_3 Y_3 + \beta_5 Y_5)(1 + \beta_2 Y_2) \tag{4.6}$$

where β_i are calibration coefficients, and 1 has been added to the denominator of the equation to avoid division by 0. The model also changes the value of the threshold (T) based on the input of external stressors value.

$$T = T + stress \tag{4.7}$$

Craving (Cr) and withdrawal (W) symptoms are also modelled, such that the curve is initiated with the initial nicotine use and grow over time,

$$W = d_3 Y_3 (T - Y_1)(Y_{0w} + Y_1) \tag{4.8}$$
$$Cr = d_5 Y_5 (T - Y_1)(Y_{0c} + Y_1) \tag{4.9}$$

where d_3, d_5, Y_{0w} and Y_{0c} are calibration coefficients.

Figure 4.1 shows the output of the CTMoS for three days as described by Bobashev et al. (2017). It can be observed in the figure that after each smoking event, the nicotine level peaks and that this is then followed by a decrease in nicotine level until the next smoking event. On the other hand, craving and withdrawal symptoms are increasing when the smoker is not smoking over a longer period (e.g. while sleeping).

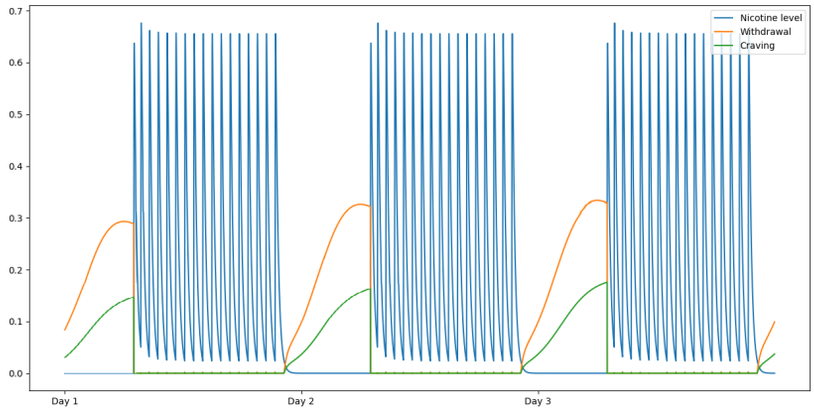


Figure 4.2: The output of the CTMoS for three days, the output is based on Bobashev et al. (2017)

Figure 4.3 shows 20 days of output of the CTMoS after 20 years of smoking, with 14 days of smoking absence, followed by relapse, as described by Bobashev et al. (2017). The figure shows the effect of smoking absent on the craving and withdrawal curves. While the withdrawal symptoms curve increases at the start, and then gradually decreases over a couple of days until it vanishes, the craving level curve increases at the beginning, and remains largely stable and high until the relapse.

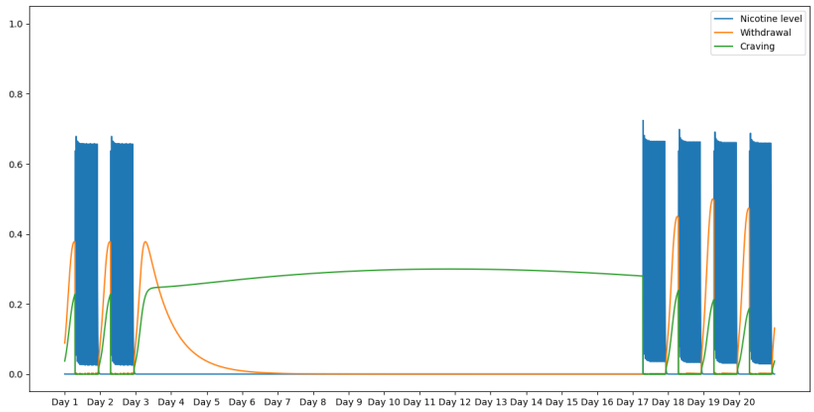
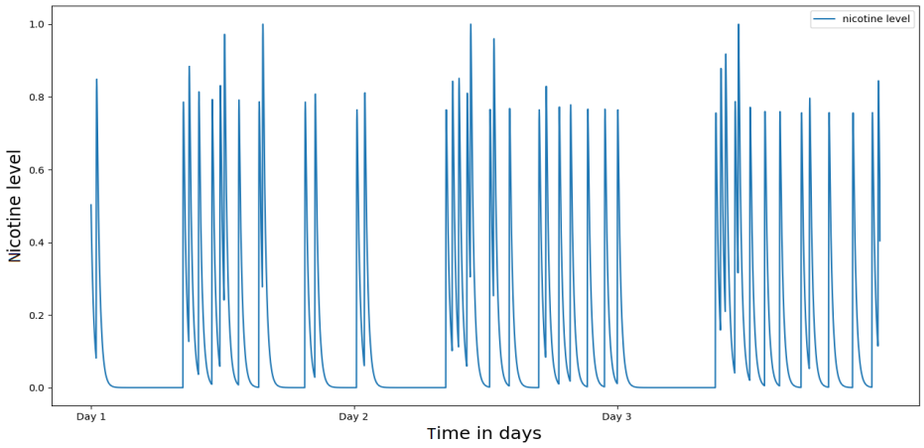
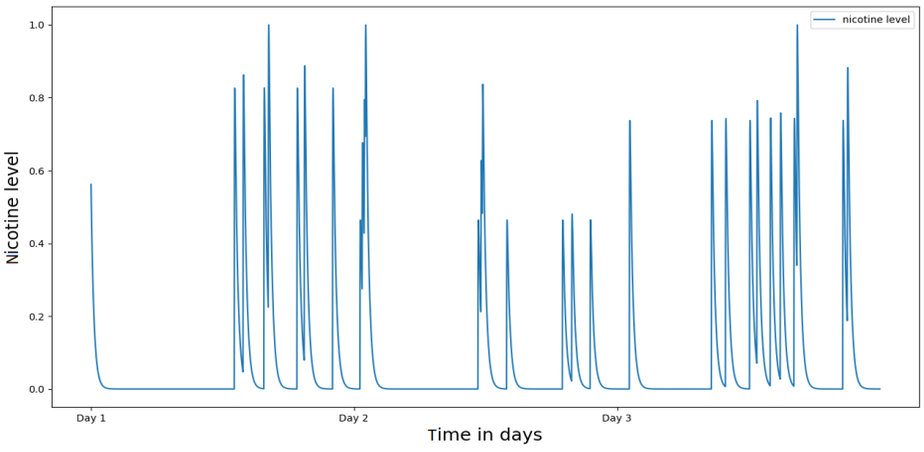


Figure 4.3: A 20 days output of the CTMoS after 20 years of smoking, with 14 days of smoking absence base on Bobashev et al. (2017)

Although CTMoS can simulate a general long-term change in smoker's behaviour (e.g. the change in craving and withdrawal symptoms during smoking absent periods), it lacks the ability to reflect the individual differences between smokers' behaviour. Figure 4.4 describes the smoking behaviour of two real smokers over three days. It is clear that individual data modelled using CTMoS results in different patterns.



(a) Example 1: Three consecutive days randomly selected from a participant.



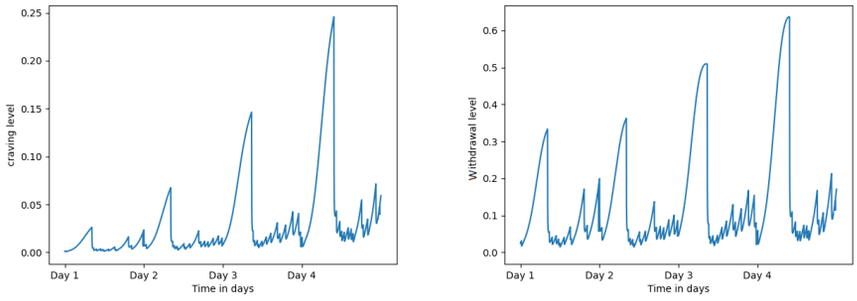
(b) Example 2: Three consecutive days randomly selected from a participant.

Figure 4.4: Two examples of three consecutive days randomly selected from two different participants (Bobashev et al., 2017).

Given that in the CTMoS, at each time-step, each output-process is relying on the output from the previous time-step, the last smoking event in the first day of data collection for each participant was used as a set-point to initialise all parameters in the CTMoS. Once this was set, the data generated for that day was eliminated from the dataset.

As explained before, CTMoS is simulating the long-term change over years, and this in turn causes a gradual increase in craving and withdrawal values over the days as illustrated in Figure 4.5. However, since this research is not targeting the long-term

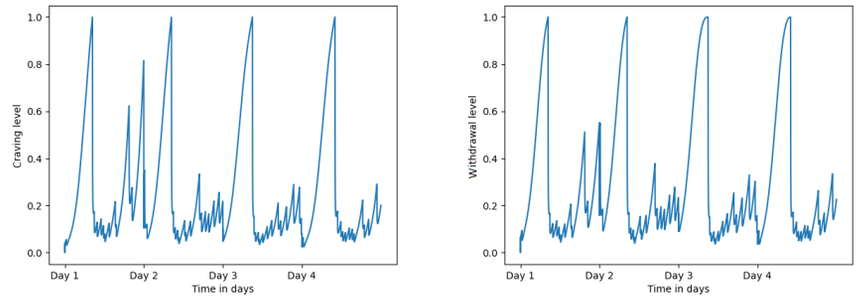
change in the smoker's behaviour, the craving and withdrawal values were normalised between 0 and 1 for each day, as illustrated in Figure4.6



(a) Example1 of craving level for four consecutive days.

(b) Example2 of withdrawal level for four consecutive days.

Figure 4.5: The craving and withdrawal output based on the original CTMoS.



(a) Example1 of craving level for four consecutive days.

(b) Example2 of withdrawal level for four consecutive days.

Figure 4.6: The craving and withdrawal output after normalization.

4.3 Relation between CTMoS nicotine craving and cue-induced craving

While CTMoS has the ability to simulate the internal level of nicotine in the human body, and the smoker's craving for nicotine, this data is not equivalent to validated lab tests. Given that it is impossible to continuously carry out lab test for research that targets the development of smoking cessation app, we require a validated model

that can be used for ML based apps, for automatic prediction of smoking events and craving level value. This type of research requires continues data collected over a long period (e.g., data collected at a frequency of one sample per minute, for at least two weeks).

Therefore, in order to validate the link between external triggers and craving level, and the behaviour of the craving curve over time, QSense dataset (Naughton et al., 2016) was used, as it offers the possibility to test the reliability of CTMoS in predicting craving level, by comparing the calculated smoking craving to an actual craving level as reported by smokers alongside their smoking events. In contrast, the Smartphone Sensors data set described in Section 4.4, does not contain reported smoking craving level, as we wished to minimize engagement with the app, in order to capture as naturalistic data as possible.

The following analysis aimed at validating the CTMoS using human participant data, as apposed to animal- lab data used for the model's development. Data collected from 41 participants was used (Naughton et al., 2016). However, most of the participants did not commit to the whole 14 days of pre-quitting data collection, and as such, the data from only 17 participants was used in this research. Data from 11 participants was excluded because they reported less than three days in total, which makes it impossible to extract the smoking pattern from this short period. Twelve participants were excluded because they reported less than 2-smoking events in more than half the pre-quitting period, meaning they did not satisfy our definition of 'smoker' (minimum 5 cigarettes per day). A problem was noted in the dataset, where the data collection days for 8 participants were not consecutive. Therefore 1-participant was eliminated because he/she had only 3-days with more than 2-days gap between them. For the other 7 participants the largest consecutive days (more than 2 days) that had no gap period were kept from their data (i.e. gap period is missing data for more than one day). The final data set contains 128 days, and 835 total reported smoking events (M=6.52, SD=3.87 per day). Participants demographics are shown in Table 4.2.

In the QSense study, smokers not only reported each time they smoked, but they also completed a short assessment following each reported smoking event (see section 4.1) for reminder). Before data analysis, the craving, stress and depress values were normalized to each participant, Data normalization was performed in two ways. First, craving level was normalized to each participant for the range from (0-5), and the range (0-4) for both stress and depress rating, such that the lowest level reported was set to 0, and the highest level was set to 4. Second, data were normalized to a binary data,

Table 4.2: QSense Participant characteristics N =17

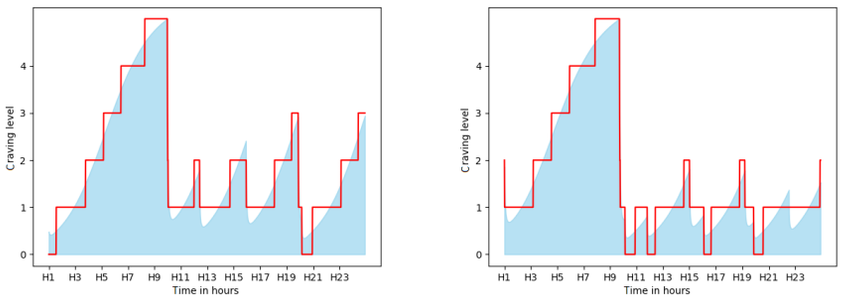
Gender	Male:8, female: 9
Age group	<20 years :0 (20 - 30) years:5 (31 - 40) years :2 (41 - 50) years:5 >50 years: 4
Cigarettes per day	<10: 5 11-15:2 >15 :10
Live with other smokers	No:9 Yes:8
Difficult Situation	When socializing:4 First thing in the morning:6 When angry or stressed:5 During an urge: 2

such that for each participant, the craving levels of 0, 1, and 2 were considered 'low craving', while rating 3, 4, and 5 were considered high craving. This was done to overcome the individual subjectivity in reported values, and to normalise the different scales used for different measures.

The time stamp related to the collected smoking events was used for calculating the craving level using CTMoS equation 4.9. This represented the CTMoS craving value, as illustrated in Figure 4.7. The CTMoS craving level was normalized (0 - 5) and rounded to the nearest integer to reflect the QSense craving scale.

First, we tested the CTMoS, assumption which suggests that long absence from smoking can lead to an increase in the craving level. This assumption had been further supported by other research (Chandra et al., 2011). Using QSense dataset out of the 121 days that were used for this analysis (7 days were excluded because the smoker reported only one smoking event in the whole day) in 62 days (51.24%), the first reported smoking event in the day had a craving level which was equal to the highest reported craving level over that day, and 14.88% of the days had flat reported craving level throughout the day. This is compared to only 33.88% of days where the craving level reported for the first cigarette of the day was lower than other cravings reported later in the day.

The reason for this 33.88% can be attributed to factors that may increase smokers craving such as stress or depression levels. In order to investigate this assumption, we compared two situations; first, when the first cigarette had the highest reported craving



(a) Example1. (b) Example2.

Figure 4.7: Example from two days, shows CTMoS craving level using QSense dataset, shaded blue area is normalized (0-5) craving, the red line is the rounded craving.

value, and second, when other smoking events had the highest craving value in the day. As shown in Table 4.3.

Out of the 128 day only the times when the participant completed the cues and craving levels were used for the analysis (i.e. when the smoker touched the reporting button in the app), meaning only 118 days. It should further be noticed that the most frequently selected value in each cue was usually the first choice (e.g. stress or depress level equal 0) in all cues , therefor to have a better understanding into what influences craving levels other than absence of nicotine, the most frequent and the second most frequent reported value from the assessment are reported separately in the table. Of the 118 used days, 51 were when the first smoking event had the highest craving ratings and in 35-day other smoking events had the highest craving ratings.

Table 4.3: The difference between when the first cigarette has the highest craving value and when other smoking events have the highest.

		Stress level	Depress level	Current situation	The presence of other smokers	Presence of people
The most frequent reported value	First smoking event	0	0	1 (‘Home’)	1 (‘None’)	1 (‘None’)
	Another smoking event	0	0	1 (‘Home’)	1 (‘None’)	1 (‘None’)
The second most frequent reported value	First smoking event	2	**2**	**4 (‘Other’)**	**2 (‘With other smokers’)**	2 (‘Friends’)
	Another smoking event	2	**4**	**2 (‘At work’)**	**3 (‘With other no smokers’)**	2 (‘Friends’)

As can be seen from the table, while both conditions (first smoke have highest craving rating, and other smoking events have highest craving rating) have instances of both the most frequent and the second most frequent value, the second most frequent value is different in term of depress level, current situation, and presence of other people cue. When smokers reported higher craving level in events that are not the first cigarette of the day, they tended to report higher depression level, being at work, or in the presence of other who are not smoking.

To measure the accuracy of CTMoS, different error rate measures were calculated to examine the difference between the rescaled craving level and the calculated craving value using CTMoS. Table 4.4 shows the average Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Normalised Root Mean Squared Error (NRMSE). These measures are commonly used to measure the average square difference between the actual and the estimated value (ignoring the times when the participant did not complete the cue or the craving buttons).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{observed} - y_{model}\right)^2 \tag{4.10}$$

$$RMSE = \sqrt{MSE} \tag{4.11}$$

$$NRMSE = \frac{RMSE}{y_{observed\ max} - y_{observed\ \ min}} \tag{4.12}$$

Table 4.4: All participants, MSE, RMSE, and NRMSE of the predicted nicotine craving.

Participant number	For rescaled craving level			Binary low-high craving level		
	MSE	RMSE	NRMSE	MSE	RMSE	NRMSE
1	6.194	2.489	0.622	0.371	0.609	0.609
2	7.547	2.747	0.549	0.194	0.44	0.44
3	3.576	1.891	0.63	0.474	0.688	0.688
4	2.775	1.666	0.555	0.308	0.555	0.555
5	4.194	2.048	0.512	0.276	0.526	0.526
6	3.881	1.97	0.493	0.513	0.716	0.716
7	3.223	1.795	0.359	0.247	0.497	0.497
8	5.972	2.444	0.489	0.538	0.734	0.734
9	3.845	1.961	0.392	0.442	0.665	0.665
10	2.88	1.697	0.339	0.207	0.455	0.455
11	2.87	1.694	0.339	0.245	0.495	0.495
12	2.02	1.421	0.474	0.5	0.707	0.707
13	1.151	1.073	0.268	0.087	0.295	0.295
14	1.142	1.069	0.534	0.286	0.535	0.535
15	2.762	1.662	0.831	0.308	0.555	0.555
16	4.594	2.143	1.072	0.435	0.659	0.659
17	2.934	1.713	0.571	0.614	0.784	0.784
Average	**3.67**	**1.87**	**0.53**	**0.34**	**0.57**	**0.57**

From the above table, CTMoS has an overall average error equals to 0.53, and the average error of predicting the low-high craving level is equals to 0.57.

Table 4.5 shows the accuracy score of predicting each craving level (ignoring times when these were not reported). First, the error per craving level was calculated to the exact craving level, and then with 20% accepted error around the calculated craving value.

Table 4.5: The accuracy score of predicting each craving level.

Craving level	**0**	**1**	**2**	**3**	**4**	**5**
craving level accuracy score	23%	62%	9%	10%	7%	36%
craving level accuracy score with (20%) accepted error	71%	89%	63%	36%	46%	39%

Table 4.6 shows the accuracy of predicting binary low-high craving level. In other words, this presents to what extent CTMoS can predict whether a smoker is having a low craving, or whtheer they are experiencing high craving level for smoking.

Table 4.6: The accuracy score of predicting low-high craving level.

	Low smoking craving	**High smoking craving**
Prediction accuracy	84%	42%

While CTMoS was able to capture the craving level to some extent, it still has high level of missprediction in some places. Bobashev et al. (2017) stated that the craving value could be affected by external stressors (Figure 4.1 and Equ. 4.7). It is expected that the difference between the calculated craving and the real craving value could be subject to other cues that have not been modelled in the Qsense data. The relation between accuracy of CTMoS and reported use can be seen in table Table 4.7(a) for the (0-5) craving level and Table 4.7(b) for the binary-level craving.

Table 4.7: The relation between external cues and the ability of the CTMoS in predicting smoking craving levels, and low-high craving level.

(a) Craving level range from (0-5)						
		Stress level	Depress level	Current situation	The presence of other smokers	Presence of people
The most frequent reported value	Correctly predicted	0	0	1 (‘Home’)	1 (‘None’)	1 (‘None’)
	Missed predicted	0	0	1 (‘Home’)	1 (‘None’)	1 (‘None’)
The second most frequent reported value	Correctly predicted	2	**1**	2	3	2
	Missed predicted	2	**2**	2	3	2
(b) Binary low-high craving level						
The most frequent reported value	Correctly predicted	0	0	1 (‘Home’)	1 (‘None’)	1 (‘None’)
	Missed predicted	0	0	1 (‘Home’)	1 (‘None’)	1 (‘None’)
The second most frequent reported value	Correctly predicted	2	2	2	**3 (‘with others who are not smoking’)**	2
	Missed predicted	2	2	2	**2 (‘with others who are smoking’)**	2

While the highest frequency of reported values is the same in both situations (and it is the first selection in all cues), at the second most frequent reported value, it seems that CTMoS mis-calculates the craving level when the person is more depressed. and that CTMoS mis-predicts craving strength (i.e. high or low craving) when the smoker is in the presence of other smokers.

Based on the above analysis, the CTMoS generated internal craving level can generally reflect the overall increase and decrease in smokers’ cue-induced craving over the day, and especially when the smoker experiences a low craving for smoking. However, CTMoS may mis-predict craving levels due to missing information on other external clues.

It is expected that applying CTMoS on a larger dataset can give better insight into its accuracy to reflect smoking craving level. To conclude, the CTMoS can reliably be employed in applications where the aim is to reflect the increase and decrease in the influences of internal –drivers on smoker’s behaviour.

4.4 Smartphone Sensors dataset study design

The smartphone sensors data was collected during this PhD. The study was approved by MMU Research Ethics and Governance Committee, EthOS Reference Number: 0441. The unique contribution of this data is that to date, there is no publicly available dataset that links continuously, real time collected data from Smokers' mobile sensors, alongside their real-time reported smoking events, for two weeks. For example, the QSense data only has EMA cue-induced reported craving factors. This was sufficient to validate the CTMoS, but Naughton et al. (2016) could not share GPS data due to ethical limitation, and it did not collect any other sensor data. Therefore QSense data was insufficient to develop and evaluate the models proposed in this PhD.

Participants in the study were adult smokers (minimum 5 cigarettes per day for minimum 6 months), over 18 years old, with a good level of English so that they could provide informed consent and understand the study instructions. Participants had to use an Android mobile phone. Participants were excluded if they were under 18, self-reported to have health issues that could affect their participation, did not use an Android phone, had difficulties in reading or speaking English or suffered from a disability that prevented them from using mobile devices.

When a participant expressed an interest in taking part in the research, they were provided with an information sheet and then asked to sign a consent form. They were then asked to complete a short survey regarding demographics (i.e. their gender, age, number of cigarette they smoke per day, and years of smoking).

Next, the data collection app was installed on the participant's mobile phone. During this initial setup of the app, participants were explained how to report the smoking event. The data collection period continued over a 2-weeks period, during which period there were no restrictions on the participants' daily behaviour, as the point of the study was to capture the real-life behaviour of the smoker in their natural environment. The only thing requested from the participants was to keep the GPS on and report every cigarette they smoked in real-time. Once the 2-weeks period was over, the participant returned to the lab, the data was downloaded from their phone, the participants were debriefed and receive a compensation voucher (£10 shopping voucher) as a thank-you for their time and effort in helping in this research.

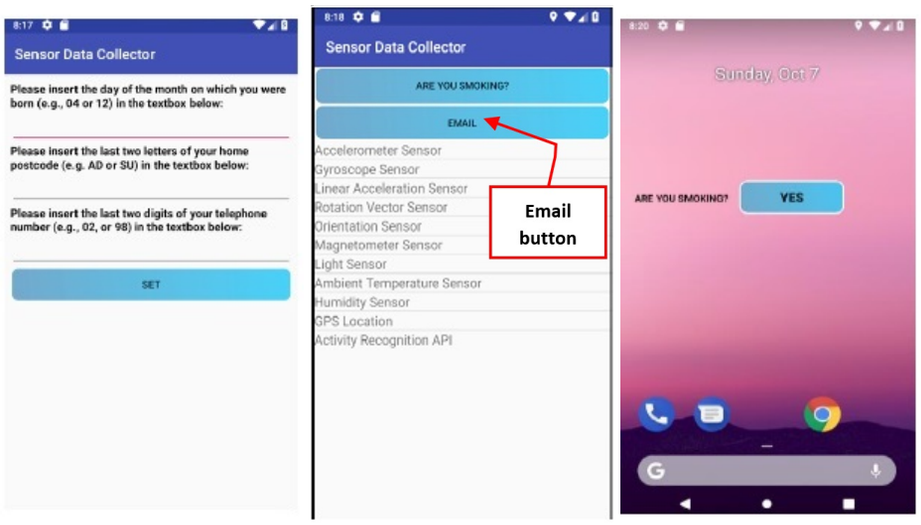
The Android phone application was developed in Java using the Android Studio (IDE). The goal of the developed app was to collect as much environmental data as possible with minimum engagement from the participant.

Figure 4.8(a) shows the first screen that appears when the app is first installed, it

generate a unique user ID that is then used to anonymise the participant data. This interface is only displayed when the app is used for the first time, and is never shown again during the use of the app.

The user interface (UI) shown in 4.8(b) is simple and user-friendly. Importantly, the interface was designed to provide no feedback to users, to avoid any influence on the participants' behaviour, as providing feedback on behaviour is a well-known behaviour change technique (Michie et al., 2013). The app enabled participants to report smoking events using two ways, either by using a simple single button inside the app 4.8(b), or by pressing a Widget on the home screen of the smartphone as can be seen in 4.8(c). The app interface in 4.8(b) also shows a list of sensors, from which the app automatically collects data from in the background.

The email button (4.8(b)) in the app is used to enable the sharing of dataset file. When the button is pressed, the user can either send the dataset file by email or send directly to external device using a Bluetooth connection.



(a) Participant registration screen. (b) Application main layout. (c) Smartphone home screen.

Figure 4.8: Mobile application User Interface (UI) (Abo-Tabik et al., 2019, 2020)

All complex app functions were kept in the background while the interface was kept simple; this means that the user does not need to initiate the start/stop of the data collection. Furthermore, even if the user turns-off the app (intentionally or accidentally), the app will continue to passively collect data using an Android background service function. The background service collected samples (one sample per minute)

using built-in sensors (accelerometer, gyroscope, light, see Table 4.9 for details). In order to improve mobile device performance and prolong battery life, Android devices place restrictions on running background services. The developed app background service was programmed to restart itself whenever it is terminated by the OS. This was required in order to keep the service running in the background, to enable continuous data collection over the two-weeks period. Table 4.8 lists the precise features, and their description.

Table 4.8: Smartphone Sensors app functions description.

Function	**Type**	**Description**
Registration screen	Form	Used to generate the participant unique ID.
Are you smoking	Button	Used to report smoking events.
Email	Button	Used to share and transfer collected data.
Are you smoking (YES)	widget	Used to report smoking events without opining the app
Background service function	Android service	Used for passive data collection in the background.

All collected data, along with labels of the smoking events, were stored on an internal SQLite database. The point at this stage of the research was to collect as much data as possible, to identify correlations between smoking events and the sensors readings, and as such, sending this volume of data over wi-fi was proving too costly, with potential risk for data-loss. Table 4.9 details the types of collected data.

To ensure that the designed data collection app achieved ease of use and efficiency in data collection, each participant completed a short survey to report on their experience using the app following the two weeks data collection period. Three 5-points Likert scales questions were used. In the first set responses ranged from (1)Very Poor, (2) Poor, (3) Neutral, (4) Good, and (5) Very Good); In the second set responses ranged from: (1) very problematic, (2) problematic, (3) slightly problematic, (4) neutral, and (5) no such problem) , and the third category questions related to reporting smoking behaviour, and used the following values: (1) Strongly agree, (2)Somewhat agree, (3) Neither agree nor disagree, (4) Somewhat disagree, (5)Strongly disagree.).

Table 4.10 shows the survey outcome as related to the design and functionality of the data collection app. In general, the app has very good feedback from all 5 participants.

Table 4.9: The Smartphone Sensors dataset data dictionary (Abo-Tabik et al., 2019, 2020)

Name	Type	Collected data	Description
ID	String	One time input string	This is a unique ID that Identifies the user data;
Timing value	Time stamp	DD-MM-YYYY, HH MM: SS	Timestamp auto generated every minute
Accelerometer	Motion sensors (float)	Three x,y,z raw values	Passively collected sensor data every minute
Gyroscope	Motion sensors (float)	Three x,y,z raw values	Passively collected sensor data every minute
Linear acceleration	Motion sensors (float)	Three x,y,z raw values	Passively collected sensor data every minute
Rotation vector	Motion sensors (float)	Three raw rotation vectors along the x-,y-,z-axis.	Passively collected sensor data every minute
The magnetic field	Environmental sensors (float)	Three raw ambient magnetic field in x-,y-,z- axis.	Passively collected sensor data every minute
Light level	Environmental sensors (float)	One raw value.	Passively collected sensor data every minute
Ambient temperature	Environmental sensors (float)	One raw value.	passively collected sensor data every minute
Relative humidity	Environmental sensors (float)	One raw value.	Passively collected sensor data every minute
GPS location	Location value (float)	Three coordinates latitude, longitude, and altitude.	Passively collected every minute
Activity	String	Google activity recognition API (Still, Running, Walking, Cycling, Tilting, and Driving).	Passively collected every minute
Smoking labels	Integer	boolean (0 and 1)	The app record 0 every minute except when the smoker report smoking event the app record 1

Table 4.10: The Smartphone Sensors app design and functionality survey.

Question	**Responses**
First group of questions	
Ease of use	very good = 3 good = 2
Functionality	very good = 1 good = 3 Neutral =1
Speed	very good = 4 good = 1
Content	good = 4 Neutral = 1
Look and feel	good = 3 Neutral = 2
Second group of questions	
I experienced bugs	No such problems = 3 Slightly problematic = 2
The app was visually unappealing	Not problematic = 4 Slightly problematic = 1
The app was confusing to use	No such problems = 4 Not problematic = 1
The app was missing features I needed	No such problems = 4 Not problematic = 1
The app crashed	No such problems =3 Problematic = 1 Not problematic = 1
The app drained battery life	Problematic = 1 Very problematic =1 Slightly problematic = 2 No such problems =1
Third group of questions	
Using the App did not make me more aware of my smoking behaviour	Somewhat disagree = 4 Somewhat agree = 1
Having to report my smoking events, made me think about how many cigarettes I smoke	Strongly disagree = 1 Somewhat agree = 1 Strongly agree = 2
Using the App made me feel embarrassed about my smoking habit	Strongly disagree = 2 Somewhat disagree = 1 Neither agree nor disagree = 1 Somewhat agree = 1
Using the App made me aware of when I am more likely to smoke	Neither agree nor disagree = 2 Somewhat agree = 3
Using the App made me think that I should try to stop smoking	Somewhat disagree = 1 Neither agree nor disagree = 3 Somewhat agree = 4

4.5 Smartphone Sensors dataset processing for ML modelling

This section explains the structure of the data set that was used for modelling smokers' behaviour, and automatic smoking events prediction using ML methods. The data was collected using the study design reported in section 4.4, and described in (Abo-Tabik et al., 2020).

This project aimed to design a model that is able to predict smoking event using passively collected data from smartphone sensors as an input to the prediction model, and combine it with a mathematical model of internal factors that influence smoking behaviour, in particular nicotine level. Therefor the aim of collecting the Smartphone Sensors dataset was to investigate what type of automatically collected data from mobile devices is most useful for modelling smokers' daily behaviour and smoking events prediction. The goal was to see whether such data could be effectively collected and used, in order to reduce the need for manually entering data by participants, a method previously identified as highly unreliable (Businelle et al., 2016; Naughton et al., 2016; Schick et al., 2018).

In this dataset, the data were collected using the app described in section 4.4, and involved data from several sensors, as explained in Table 4.9. Six participants met the inclusion criteria for the smartphone sensors study. One of the 6 participants was later excluded from data analysis due to lack of engagement (only 2-days of data were collected). Table 4.11 reports demographic information for the 5 participants. Although the number of participants appears small, the study by Schick et al. (2018) modelled smoking behaviour using 4 participants. Hence 5 participants were a sufficient number to model smoking behaviour. Besides, the Smartphone Sensors dataset contains a total of 64 days (13 days from four participant, and 12 days from one participant), involving 1440 sample per day (totalling 92,160 samples in sum from all participants), with 523 smoking-events reported by the 5 participants (M = 7.92, SD = 4.78), making the dataset sufficient large for modelling a ML problem.

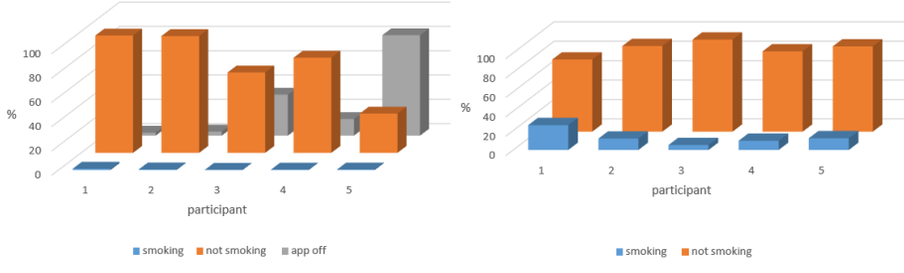
As built-in sensors availability is varied between different mobile models, the project had to use data only from sensors that were available on all devices. Therefore, only accelerometer sensor readings were used, along with location data.

The smoking events were labelled to reflect three scenarios: (0) not-smoking, (1) smoking and (2) app-off (representing gaps in the dataset due to, for example, participant's mobile phone being off).

Table 4.11: Smartphone Sensors dataset Participant characteristics N =5

Gender	Male:2, female: 3
Age group	<20 years :0 (20 - 30) years:2 (31 - 40) years :1 (41 - 50) years:2 >50 years: 0
Cigarettes per day	<10: 3 11-15:1 >15 :1
Years of smoking	<10 years: 1 (10 – 20) years:2 >20 years:2

Figure 4.9(a) shows the frequency of each class for all five participants. It is clear that these classes are unbalanced, as the number of not-smoking events is much more frequent than the two other classes. Where out of the 1440 daily samples, there are around an average of 8 smoking events per day and the rest are either not-smoking or app-off events, this problem can later affect the learning process of the ML model. To avoid this limitation, each 1-min smoking event was converted to a 30-min window followed the smoking event, hence reducing the ratio of smoking to not-smoking events. In this case, the prediction model could learn to predict the time of the day when the smoker is most likely having a cigarette, based on the external predictors associated with their behaviour. Furthermore, it is assumed that app-off is a not-smoking event to remain cautious. Figure 4.9(b) shows the frequency of events for each of the five participants after applying these changes.



(a) Percentage frequency of the three labelling categories.

(b) Percentage frequency after processing the data.

Figure 4.9: Data set pre-processing (Abo-Tabik et al., 2020).

For the final step in the design process of the Smartphone Sensors dataset, CTMoS

output was calculated and added to the dataset. To do so, the original reported smoking events were used as input to the CTMoS in order to calculate the nicotine, craving, and withdrawal symptoms levels and threshold value during a 13-day period. The missing 24 hour period is made of two half-days, one at the start (when the participant had their app installed) and the other at the final day of the data collection period (when the participant returned the data set to the lab). The calculated CTMoS data (i.e. nicotine level, craving, withdrawal, and threshold) along with collected data (e.g., Accelerometer (x, y, and z) values, GPS Location, etc.) and the labels were combined to create the dataset for each participant. Figure 4.10 illustrates the process of data base creation

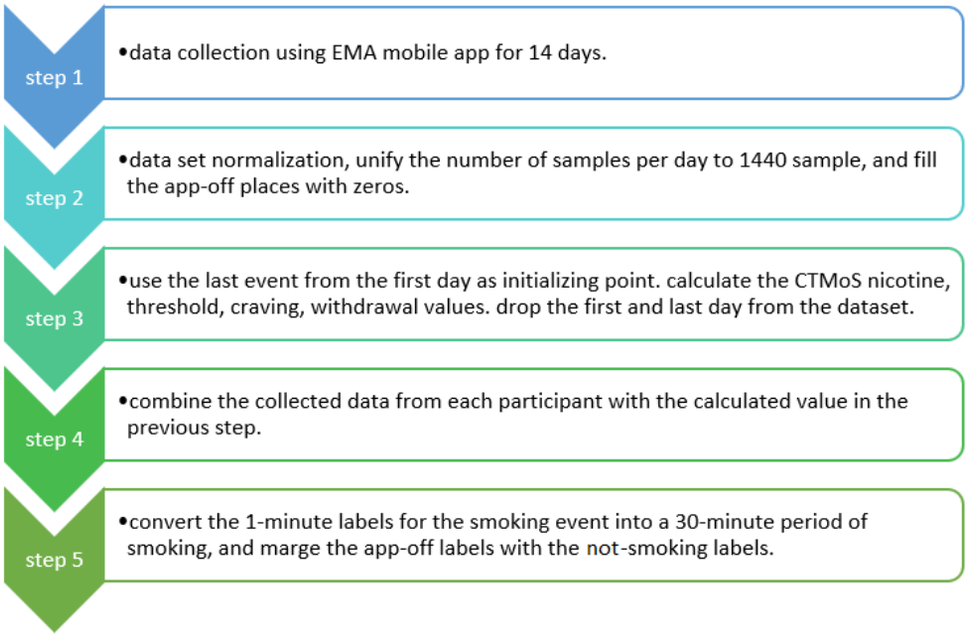


Figure 4.10: Overview of the study: Data collection and processing steps for each participant in the Smartphone Sensors dataset (Abo-Tabik et al., 2020)

4.6 Conclusions

This chapter describes the CTMoS; it shows that this model can be used to simulate internal smoking craving factors which are nicotine level in the blood, nicotine craving and withdrawal symptoms. The chapter highlights the ability of CTMoS to reflect the overall increase and decrease in nicotine craving, and especially when the smoker

experiences a low craving for smoking, and also show how it mispredicts high nicotine craving levels due to missing information on other external clues. The chapter validated the CTMoS with human participants data collected from 17 participants using QSense study Naughton et al., 2016. The chapter also describes the Smartphone Sensors dataset, this data will be used later in designing an automatic smoking event prediction models in chapter 5 and 6. The chapter gives a detailed description of the Smartphone Sensors study design, the number of participants, and also explain the design of the mobile app that was used for data collection of this study.

Chapter 5

A 1D-CNN Model Predicting Smoking Behaviour

This chapter will present an approach for modelling smokers' behaviour to enable the automatic prediction of smoking events. The model combines CTMoS, which models internal factors that influence smoking (e.g. nicotine level), with 1D-CNN, which extracts feature patterns from mobile-phone sensors data representing external factors that influence smoking (e.g. motion and location).

As such, the combined model can adapt to individual differences in the patterns of smoking between different smokers. The model was tested on data from 5 smokers, drawn from the Smartphone Sensors dataset as described in Chapter 4.

The model was evaluated using k-fold cross-validation for each participant individually. At each round for each participant, one day was held for testing, and the rest were used for training. The model accuracy for each participant was the average accuracy of all rounds.

The smoking behaviour prediction model consists of two parts the DL model and CTMoS. At the start, to validate the 1D-CNN, three different ML models were used (SVM, DT, and RF) to evaluate their performance. The model was also evaluated using each type of input separately, to examine their contribution: (1) only accelerometer, (2) only GPS, (3) both accelerometer and GPS. As such the experimental hypothesis to be tested will be, if the four tested ML methods are significantly different in their performance and also if 1D-CNN is better in performance than the other three ML models; in addition to test whether combining more than input feature is better in performance than using only single external environmental factor. Moreover, the designed 1D-CNN was evaluated as a sequential and multi-headed parallel model. The best performance

was observed for the multi-headed parallel model-CNN. This model captured the feature patterns from 6 raw external input data (three GPS and three accelerometer) values, in order to predict smoking events to high degree of accuracy.

Then the combined 1D-CNN with CTMoS model was evaluated based on its ability to reflect the internal smoking carving factors (nicotine level, smoking craving, withdrawal). This is in addition to the model ability to predict smoking events in 5- 15- 30- 45- min prior to the appearance of the smoking event.

The 1D-CNN proved to be better than the other three ML methods in predicting smoking events based on only external factors; and by combining it with CTMoS it can predict smoker's behaviour with average f1-score equal to 0.06, 0.14, 0.24, 0.4 for 5, 15, 30, and 60 -min respectively, prior to the appearance of the smoking event, base on both internal and external factors. Overall, the model proved its ability to predict the smoker's behaviour to a high degree when the participant is regularly engaged.

The results reported in this chapter have been published in Abo-Tabik et al. (2020), but this chapter reports extended the analysis to obtain further validation of the designed model.

5.1 The architectural design and evaluation of the 1D-CNN for smoking events prediction

The first part of the smoker's behaviour prediction model is a 1D-CNN that will predict smoking events based only on external factors. The input to the model are 3 accelerometer values (x, y, and z) and 3 GPS readings (longitude, latitude and altitude). The essence of this method is to take advantage of the CNN's ability to automatically extract features, without having to rely on feature engineering methods, as these may increase time and computation; as explained previously in Chapter 3 (section 3.5).

Each of the 6 feature vectors of the input are being passed to a separate 1D-CNN, and the output of these sub-processors are combined to get the full feature vector. Random search methods were used for hyperparameter optimization. These are the sub-sample size, convolutional layer filter size, and the fully connected layer number of neurons. The final model is shown in Figure 5.1.

At the start, each of the input vectors is being sub-sampled into samples of 10-past observations that can be denoted as $X_{i-10:i}$ for the i^{th} sample; and each of these 10-observations are being passed to a separate sub-model.

The first layer in the sub-model is a convolutional layer with a filter-size = 64; the

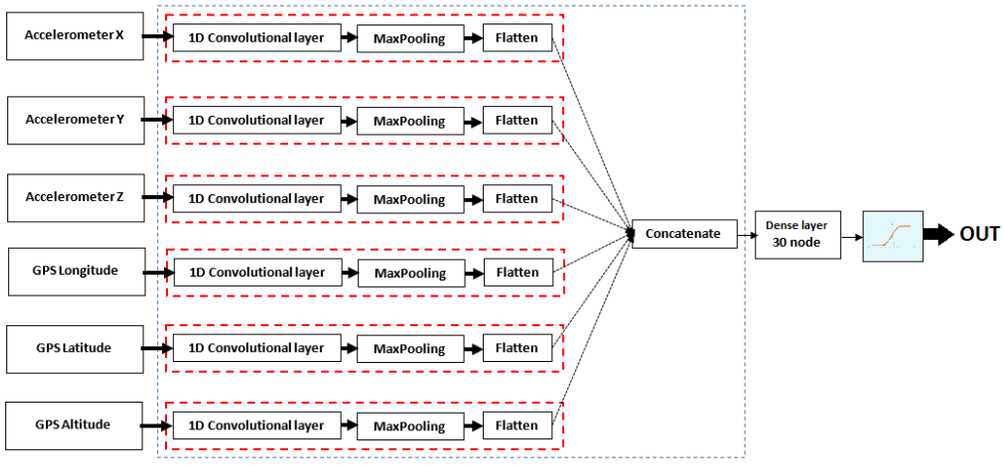


Figure 5.1: One-dimensional convolutional neural network (1D-CNN) model architecture (Abo-Tabik et al., 2020)

convolutional operation takes the form of,

$$C_i = h(W^T \bigotimes X_{i-10:i} + b_i) \tag{5.1}$$

where $\bigotimes$ is the convolution operator, W^T is the network weights, b is the bias, and h is the non-linear ReLU activation function with $l2$ weight regularisation. The output from this layer is called the feature map. Batch normalisation was added after this layer to standardise the input for each next layer. The generated feature map is then being passed to a max-pooling layer, where it merges the relatively similar features, and reduces the variance by using max function,

$$P_{i,k} = Max\,(C_{(i,k)}\,U_{3,1}) \tag{5.2}$$

where k is the filter number, and $U_{3,1}$ is sliding max window of size 3×1.

Finally, all sub-models are combined into a fully connected ANN layer, with neuron size of 30. The output layer uses the *softmax* activation function to generate a one-hot encoded output. A detailed description of the designed model can be seen in Table 5.1.

Table 5.1: A detailed description of 1D-CNN architecture.

Layer (type)	Output Shape	Param #	Connected to
input1	[(None, 10, 1)]	[(None, 10, 1)]	
input2	[(None, 10, 1)]	[(None, 10, 1)]	
input3	[(None, 10, 1)]	[(None, 10, 1)]	
input4	[(None, 10, 1)]	[(None, 10, 1)]	
input5	[(None, 10, 1)]	[(None, 10, 1)]	
input6	[(None, 10, 1)]	[(None, 10, 1)]	
Conv1 (Conv1D)	(None, 8, 64)	256	input1
Conv2 (Conv1D)	(None, 8, 64)	256	input2
Conv3 (Conv1D)	(None, 8, 64)	256	input3
Conv4 (Conv1D)	(None, 8, 64)	256	input4
Conv5 (Conv1D)	(None, 8, 64)	256	input5
Conv6 (Conv1D)	(None, 8, 64)	256	input6
batch normalization1	(None, 8, 64)	256	Conv1 (Conv1D)
batch normalization2	(None, 8, 64)	256	Conv2 (Conv1D)
batch normalization3	(None, 8, 64)	256	Conv3 (Conv1D)
batch normalization4	(None, 8, 64)	256	Conv4 (Conv1D)
batch normalization5	(None, 8, 64)	256	Conv5 (Conv1D)
batch normalization6	(None, 8, 64)	256	Conv6 (Conv1D)
max pooling1d1	(None, 4, 64)	0	batch normalization1
max pooling1d2	(None, 4, 64)	0	batch normalization2
max pooling1d3	(None, 4, 64)	0	batch normalization3
max pooling1d4	(None, 4, 64)	0	batch normalization4
max pooling1d5	(None, 4, 64)	0	batch normalization5
max pooling1d6	(None, 4, 64)	0	batch normalization6
Flatten1	(None, 256)	0	max pooling1d1
Flatten2	(None, 256)	0	max pooling1d2
Flatten3	(None, 256)	0	max pooling1d3
Flatten4	(None, 256)	0	max pooling1d4
Flatten5	(None, 256)	0	max pooling1d5
Flatten6	(None, 256)	0	max pooling1d6
concatenate (Concatenate)	(None, 1536)	0	Flatten1 Flatten2 Flatten3 Flatten4 Flatten5 Flatten6
dense (Dense)	(None, 30)	46110	concatenate
output (Dense)	(None, 2)	62	dense

5.1.1 Model evaluation

The Smartphone Sensors dataset set (described in Chapter 4) was used for model evaluation. The model was evaluated for each participant individually using the k-fold validation process. The dataset was portioned into 13 separate folds, and each fold is 1440 sample of one-day data. For each test iteration, one day will be held for testing, and the rest 12 days are used for training. This approach can help to test the generality of the model across all weekdays.

In order to minimise the chance of falling into overfitting problem; an early stop technique was used in model training. Keras offers the "EarlyStopping" function which monitors the model performance based on a validation set, and stops the training if it finds a continuous degradation in the evaluation performance or the selected performance measure (i.e. validation accuracy) stops improving during the training.

The classification model was tested to see whether the classifier could detect the smoking events based on the smoker motion using and location. The input vector, as explained in the previous section is 3 accelerometer values and 3 GPS values, and the output is a one-hot encoding that encodes (0) for not-smoking and (1) for smoking.

5.1.2 Results

In order to verify the effectiveness of the system, three other ML models were explored: SVM, DT, and RF, along with the 1D-CNN model. Each of these models, along with the 1D-CNN, were evaluated for each participant separately. A random search approach was used for hyperparameters optimization for all three ML methods SVM, DT, and RF. As result of models optimization; for the SVM the selected hyperparameters were the kernel radial basis function (Chang et al., 2010), and the degree of the polynomial kernel function was set to 3. While for the DT, RF the Gini index (Breiman et al., 1984) was used as data separation function, and max depth set to None, meaning that the tree keeps splitting until all leaves are pure (i.e. all its data points contain the same label, or when the splitting function cannot be split any further which occurs when two identical samples result in a different output).

The average accuracy of all 13 folds (i.e. the accuracy for each fold was first calculated and then the mean of the computed values was the average accuracy) was first calculated using GPS data only (Table 5.2), then using accelerometer data only (Table 5.3), and lastly, both accelerometer and GPS data were used as input to the models (Table 5.4).

$$\text{Average accuracy} = \frac{accuracy_1 + accuracy_2 + ... + accuracy_{numofdays}}{numberofdays} \quad (5.3)$$

Table 5.2: The average accuracy based on only 3 GPS values (Abo-Tabik et al., 2020).

Calculated Accuracy Category	**SVM**	**DT**	**RF**	**1D-CNN**
Participant 1 smoking	0.01	0.40	0.36	0.01
Participant 1 not-smoking	0.98	0.70	0.77	1.00
Participant 1 accuracy	0.73	0.62	0.66	**0.74**
Participant 2 smoking	0.03	0.51	0.52	0.09
Participant 2 not-smoking	0.99	0.95	0.96	0.98
Participant 2 accuracy	0.88	0.90	**0.91**	0.87
Participant 3 smoking	0.02	0.08	0.11	0.00
Participant 3 not-smoking	0.99	0.91	0.93	1.00
Participant 3 accuracy	0.94	**0.95**	0.86	**0.95**
Participant 4 smoking	0.00	0.24	0.12	0.08
Participant 4 not-smoking	1.00	0.81	0.95	1.00
Participant 4 accuracy	**0.90**	0.88	0.86	**0.90**
Participant 5 smoking	0.00	0.25	0.06	0.21
Participant 5 not-smoking	1.00	0.97	0.89	0.97
Participant 5 accuracy	0.88	**0.97**	0.87	0.88

Table 5.3: The average accuracy based on only 3 accelerometer values (Abo-Tabik et al., 2020).

Calculated Accuracy Category	**SVM**	**DT**	**RF**	**1D-CNN**
Participant 1 smoking	0.26	0.43	0.38	0.51
Participant 1 not-smoking	0.74	0.75	0.82	0.83
Participant 1 accuracy	0.62	0.67	0.7	**0.75**
Participant 2 smoking	0.19	0.37	0.34	0.63
Participant 2 not-smoking	0.88	0.89	0.95	0.95
Participant 2 accuracy	0.80	0.83	0.88	**0.91**
Participant 3 smoking	0.06	0.08	0.06	0.01
Participant 3 not-smoking	0.95	0.94	0.98	1.00
Participant 3 accuracy	0.90	0.89	0.89	**0.95**
Participant 4 smoking	0.21	0.16	0.12	0.18
Participant 4 not-smoking	0.91	0.93	0.95	0.97
Participant 4 accuracy	0.84	0.85	0.85	**0.89**
Participant 5 smoking	0.12	0.25	0.24	0.44
Participant 5 not-smoking	0.95	0.95	0.96	0.94
Participant 5 accuracy	0.85	0.86	**0.88**	0.87

Table 5.4 shows the average classification accuracy by using 6 features (accelerometer and GPS values) as input to the model.

Table 5.4: The average accuracy based on all 6 features accelerometer and GPS values (Abo-Tabik et al., 2020).

Calculated Accuracy Category	SVM	DT	RF	1D-CNN
Participant 1 smoking	0.24	0.41	0.38	0.59
Participant 1 not-smoking	0.79	0.77	0.86	0.79
Participant 1 accuracy	0.65	0.68	**0.73**	**0.73**
Participant 2 smoking	0.04	0.50	0.54	0.64
Participant 2 not-smoking	0.87	0.92	0.96	0.94
Participant 2 accuracy	0.78	0.87	**0.91**	0.89
Participant 3 smoking	0.05	0.11	0.1	0.03
Participant 3 not-smoking	0.96	0.93	0.96	0.99
Participant 3 accuracy	0.91	0.89	0.89	**0.94**
Participant 4 smoking	0.14	0.28	0.13	0.20
Participant 4 not-smoking	0.91	0.87	0.96	0.97
Participant 4 accuracy	0.83	0.81	0.87	**0.89**
Participant 5 smoking	0.12	0.26	0.23	0.47
Participant 5 not-smoking	0.95	0.95	0.96	0.94
Participant 5 accuracy	0.85	0.86	**0.88**	**0.88**

Besides model accuracy, statistical significance testing was used to evaluate the performance of the three methods. Statistically, significant testing is a common method to compare different classifiers; it is used to find if different classifiers are significantly different in their performance. The Friedman (1937) test is a well-known and widely used technique (Satu et al., 2020; Verma and Ranga, 2020; Wu et al., 2016), which shows reliable performance in evaluating the significant differences between multiple classifiers (Demšar, 2006). The Friedman test will be used to find the significant difference between multiple classifiers.

There are two null hypotheses (H0) to be examined here, each will be tested against two significant levels, (α) 0.05 and 0.1. If the comparison generates a p-value lower than α, the difference is considered as significant. The first null hypotheses to be tested if the there is no significant difference between the performance of different classifiers (i.e 1D-CNN, SVM, DT, and RF) against the hypotheses that there is at least one classifier that is significantly different in performance from at lease one other classifier. The second null hypothesis is there will be no significant difference in the performance if both accelerometer and GPS as used as inputs to the classification model. Significant

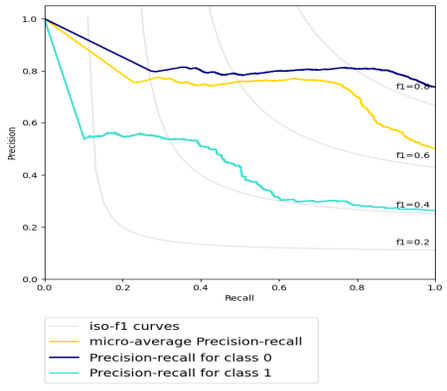
testing in this research will be applied to the model's ability in predicting smoking events.

The Friedman test was applied on the four classifiers 1D-CNN, SVM, DT, and RF, yielding $Q = 6.84, d.f. = 4, p = 0.034$. As this is less than the two levels critical levels, 0.05, and 0.1, at least one model is significantly different in its performance from the other models. Since there is data from only 5 participants; it would be hard to have an accurate post-hoc test. Therefore model accuracy metrics will be used to compare the four models performance, and by using that as shown in Table 5.4 it can see that 1D-CNN is better in performance than the other classifiers.

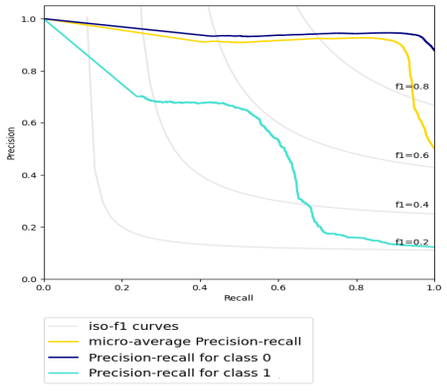
The second comparison examined the 1D-CNN with the input features of either (1) 3-accelerometer values, (2) 3-GPS values and (3) 6 input features 3-accelerometer and 3-GPS values, giving $Q = 10, d.f. = 4, p = 0.003$. This is less than α and that means the three approaches are statistically different in their performance. By comparing the results from the three Tables 5.2, 5.3, and 5.4 it can be observed that using both accelerometer and GPS values 1D-CNN as input out performs the model with either data-type separately.

It can be concluded from the results above, that both motion and location data are important to predict the smoking events, particularly when using the 1D-CNN. Additionally, in general 1D-CNN has outperformed the other three ML methods on overall accuracy between (0.73-0.94) and smoking prediction accuracy between (0.03-0.64), compared to SVM with overall accuracy between (0.65-0.91) and smoking prediction accuracy between (0.04-0.24), DT with overall accuracy between (0.68-0.89) and smoking prediction accuracy between (0.11-0.50), and RF with overall accuracy between (0.73-0.89), and smoking prediction accuracy between (0.1-0.54). It can also be noted that the performance of the 1D-CNN was low for participant 3 and 4, where the number of reported smoking events was low, which makes it hard for the model to extract behavioural patterns.

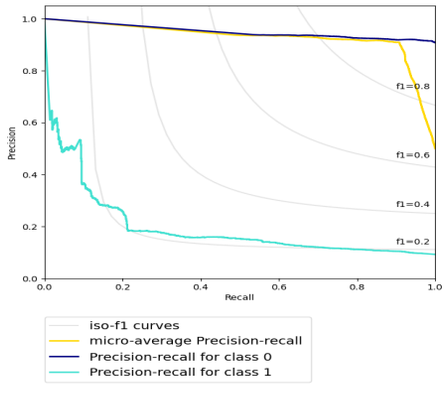
Figure 5.2 shows the Precision-Recall for each participant; it can be seen that the continuity of the data is an important factor that affects the prediction of smoking events. Despite this, the overall classification accuracy remains high (M= 0.87, SD= 0.08).



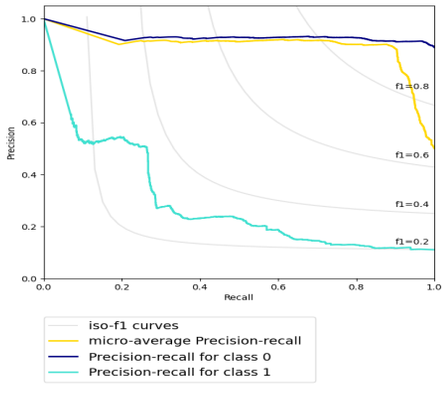
(a) Participant 1.



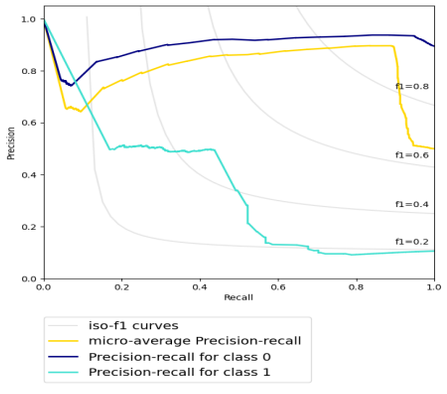
(b) Participant 2.



(c) Participant 3.



(d) Participant 4.

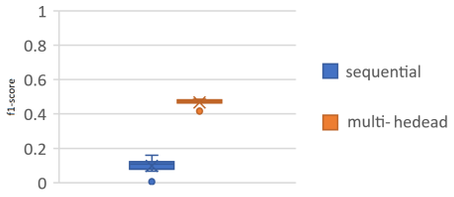


(e) Participant 5.

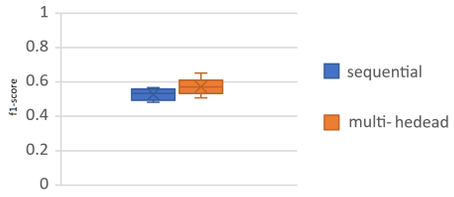
Figure 5.2: 1D-CNN model Precision-Recall.

The designed 1D-CNN model consists of six parallel sub-models that are combined to generate final 'out'. This design approach can be referred to as a multi-headed parallel design. One advantage of this approach is that it can improve the model performance by processing each of the feature vectors separately. The other advantage is that it provides design flexibility, as the model can easily be improved in the future and different configurations for each sub-model can be tested.

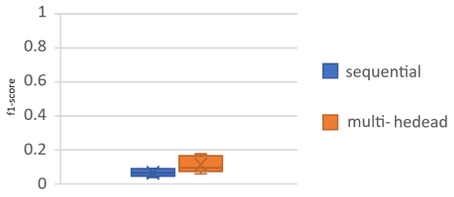
To demonstrate the efficiency of the multi-headed 1D-CNN over the sequential model(i.e. the model is designed as a stack of layers, in which each layer weights correlate to the followed layer. It takes combined input vectors and passes them to the model layers to gain the final out), Figure 5.3 shows a comparison between multi-headed 1D-CNN design and a sequential design in term of computing the f1-score for each day for each of the 5 participants.



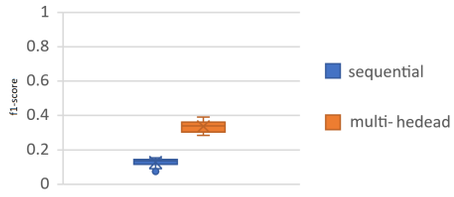
(a) Participant 1.



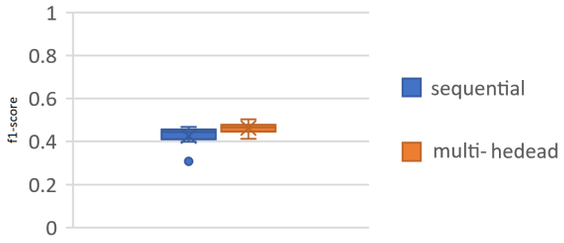
(b) Participant 2.



(c) Participant 3.



(d) Participant 4.



(e) Participant 5.

Figure 5.3: f1-score for both multi-headed model and the sequential model.

5.2 Using 1D-CNN model to predict internal smoking factors that drive behaviour

Using external factors that influence behaviour, the multi-headed 1D-CNN was shown to predict smoking events better than SVM, DT, and RF. Therefore, the 1D-CNN was chosen to be combined with CTMoS, which is used to model internal factors that influence behaviour, including nicotine, craving, and withdrawal symptoms levels. The Hybrid model for predicting smoking events should be able to make use of both internal and external factors. Figure 5.4 shows the designed DL model of smoker's behaviour.

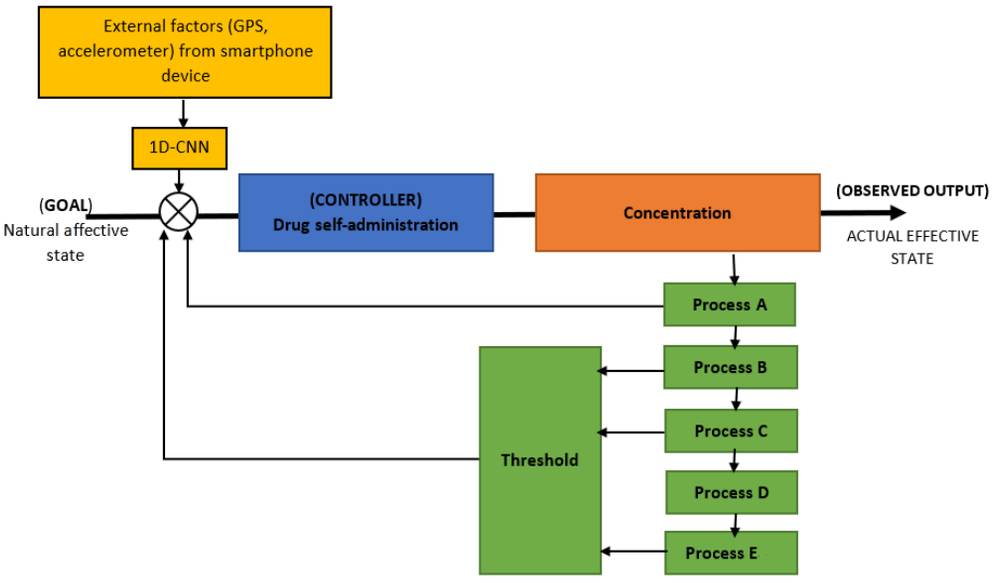


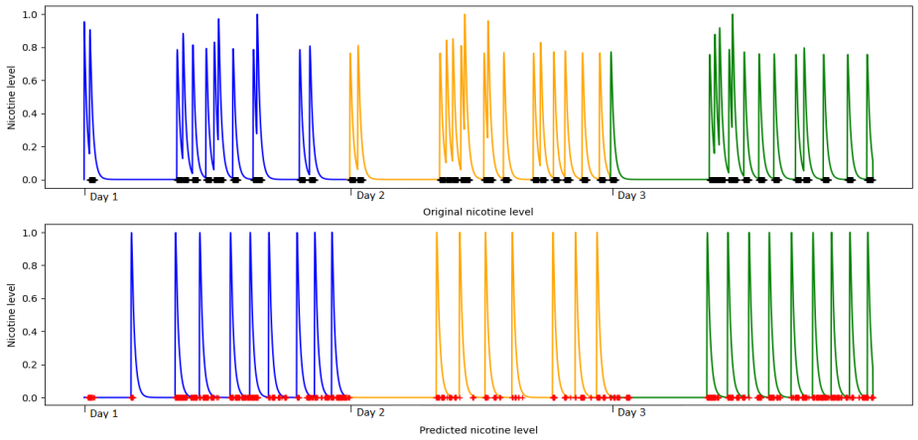
Figure 5.4: Smoking behaviour model utilising machine learning. Two predictors are used as input to the 1D-CNN model. A classification value of 1 represents a potential smoking event. This value passed to the CONTROLLER, simulating the taking of a cigarette, and re-initialising the parameters of the control model to zero (Abo-Tabik et al., 2020)

In this model, CTMoS processes A to E maintained the same functionality previously described in section 4.2. The 1D-CNN makes use of external factors (motion and location) to predict the likelihood of a smoking event. If the nicotine level is less than the threshold, then this can be an indicator of a potential smoking event. This model can be utilised for the development of smart smoking cessation apps, in order to send timely and targeted intervention messages.

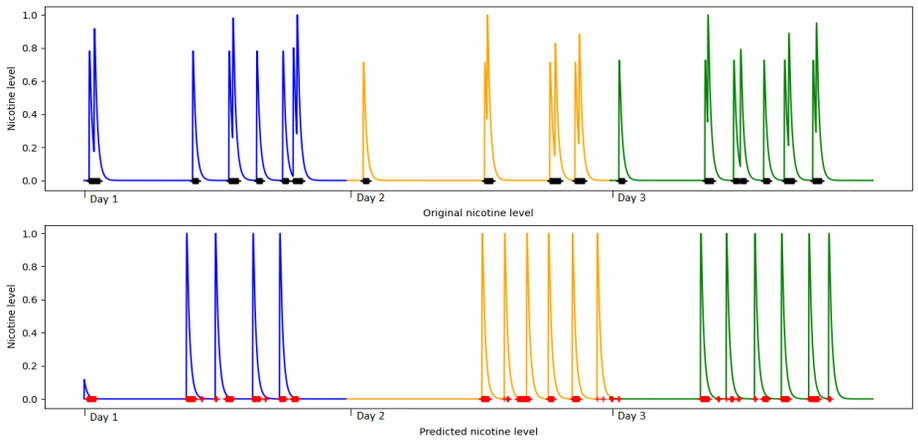
Figure 5.5 shows results of output from the combined model for 3 consecutive days from two participants. In the figure the original nicotine level was calculated by applying CTMoS on real participants smoking events; while predicted nicotine level is calculated using the model shown in Figure 5.4.

The use of 1D-CNN model as part of the CTMoS helped the inclusion of external (i.e. 6 input features from mobile sensors as input to the 1D-CNN) and internal factors (i.e nicotine level generated using CTMoS). Hence, the resultant model first uses 1D-CNN to predict smoking events using external factors within a 30-minutes window. The outcome of this is then compared to the nicotine level and threshold value as computed by CTMoS, which produces the final decision regarding the likelihood presence of a smoking event; this can help to ensure that no intervention messages are sent before the nicotine level (as derived from Equation 4.5) decreases to a level that is below the threshold (as derived from Equation 4.6).

The problem of over-sending interventions messages was previously reported by Schick et al. (2018). While from a modelling perspective this may seem a 'small error', from a health-intervention perspective this can be very harmful, as it will not only reduce the trust the user has in the app, but it may actually serve as a reminder to the fact they want a cigarette, rather than help them to overcome their urge to smoke, as reported to be the case among some participants in Schick et al. (2018). For this reason, combining the models to avoid such errors is an important improvement on previous work.



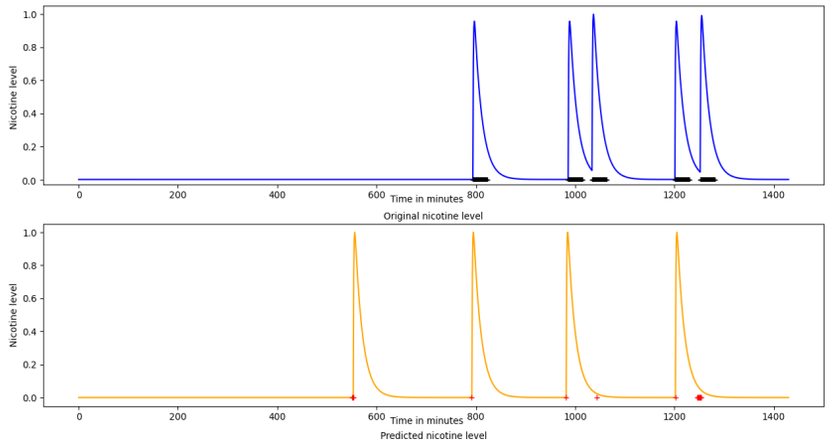
(a) Example 1: for randomly selected three day sequence from one participant.



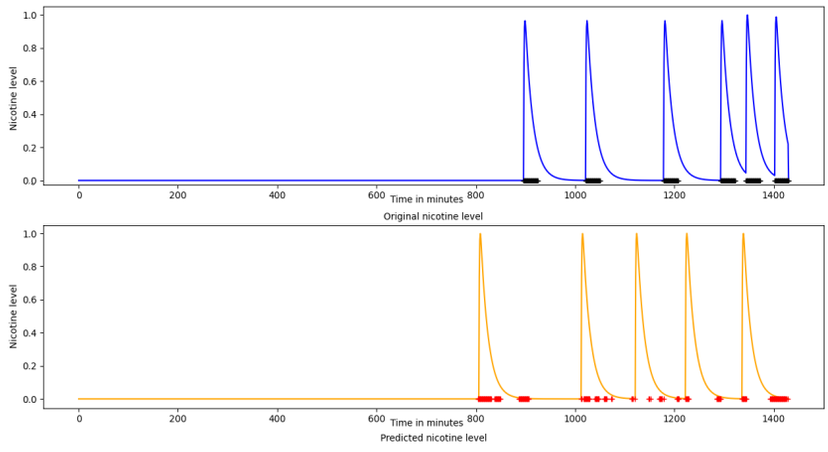
(b) Example 2: for randomly selected three day sequence from one participant.

Figure 5.5: Results of the hybrid smoking behaviour model; the output is for three randomly selected days from two different participants, each different colour present different day,the black + markers are the real reported smoking events from the original data, and the red + markers are the predicted smoking events using 1D-CNN (Abo-Tabik et al., 2020).

While the model is generally reliable in predicting smoking events, there are some errors in the predicted events. These errors could be, at least partially, be attributed to missing samples from the collected data, that may affect the performance of the 1D-CNN model. Figure 5.6 shows the predicted smoking events for randomly selected days with a high level of missed data.



(a) Example 1: for a day with high missed data.



(b) Example 2: for a day with high missed data.

Figure 5.6: Predicted nicotine level for participants with a high presence of app-off values, the black + markers are the real reported smoking events from the original data, and the red + markers are the predicted smoking events using 1D-CNN (Abo-Tabik et al., 2020).

To illustrate this, there was one participant who was reported smoking events in all days except for one. In this case, the model predicted several smoking events for that day, and it is hard to be sure whether these predictions were errorneous or whether these were predictions of unreported smoking events (Figure 5.7).

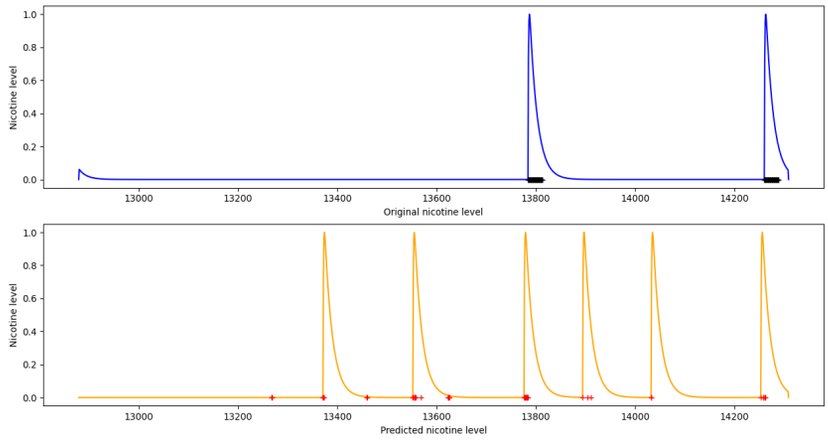
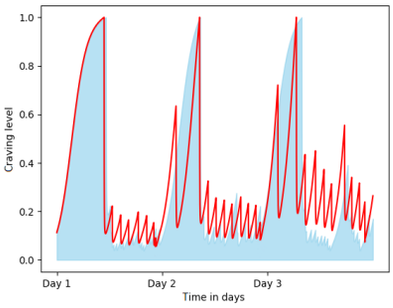
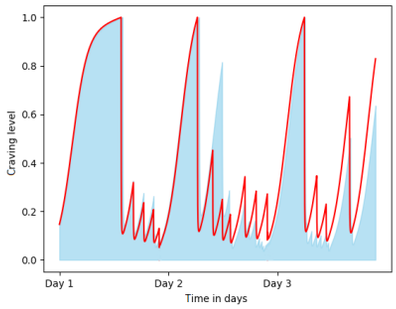


Figure 5.7: Predicted nicotine level for a committed participant with a day that has a very-low reported smoking events, the black + markers are the real reported smoking events from the original data, and the red + markers are the predicted smoking events using 1D-CNN (Abo-Tabik et al., 2020).

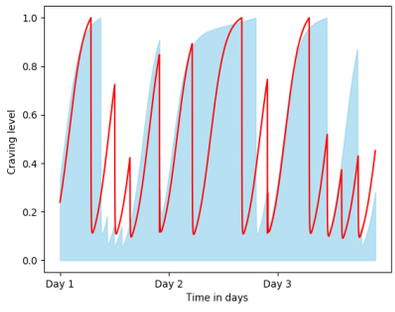
Besides nicotine level, craving and withdrawal symptoms level were also predicted using the designed model. Figure 5.8 and Figure 5.9 show the output of craving and withdrawal symptoms levels for three consecutive days from three different participants. The figures shows a comparison between the CTMoS curve as calculated using the reported smoking events from the Smartphone Sensors dataset, and the predicted craving level and withdrawal values using the designed model presented in Figure 5.4.



(a) Example 1: Three consecutive days from one participant.

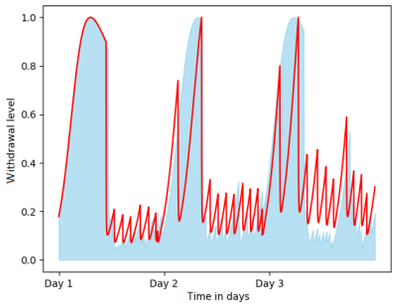


(b) Example 2: Three consecutive days from one participant.

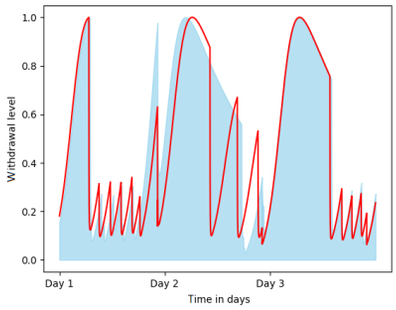


(c) Example 3: Three consecutive days from one participant, with a high level of 'app-off' values.

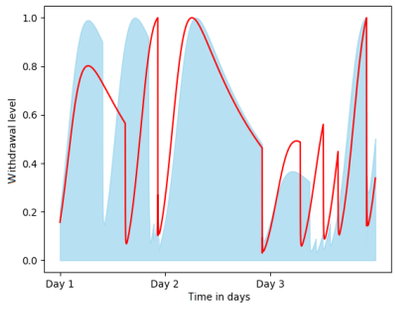
Figure 5.8: A comparison between craving levels (the blue shaded area) generated using reported smoking events, and craving level predicted using 1D-CNN and CTMoS combined model(the red line)



(a) Example 1: Three consicutive days from one particpant.



(b) Example 2: Three consicutive days from one particpant.



(c) Example 3: Three consecutive days from one participant, with a high level of 'app-off' values.

Figure 5.9: A comparison between withdrawal levels (the blue shaded area) generated using reported smoking events, and craving level predicted using 1D-CNN and CTMoS combined model (the red line)

MSE and NRMSE were used as the error criteria to measure the performance of the model on its three output: nicotine level, craving level, and withdrawal values over time. Nicotine, craving and withdrawal symptoms level curves presented in Table 5.5 were all normalised (0-1) before applying the accuracy measures.

Table 5.5: The overall error rate of the designed model of predicting nicotine level.

	Nicotine level		Craving level		Withdrawal level	
	MSE	NRMSE	MSE	NRMSE	MSE	NRMSE
Participant 1	0.07	0.27	0.03	0.18	0.04	0.2
Participant 2	0.05	0.21	0.05	0.23	0.05	0.21
Participant 3	0.02	0.15	0.11	0.33	0.09	0.3
Participant 4	0.05	0.22	0.08	0.29	0.07	0.27
Participant 5	0.03	0.17	0.02	0.15	0.03	0.16

Lastly, Table 5.6 shows the accuracy (see section 3.1, equation 3.1) of the designed model in predicting smoking behaviour and its the ability to predict smoking events based on both internal and external factors in a 5 min, 15 min, 30 min, and 1-h window prior to the occurrence of a smoking event. The table also shows the f1-score (equation 3.4) for the same time windows.

Table 5.6: The evaluation matrix for all participants, shows the prediction accuracy of the combined model.

PID	Time window	Accuracy	Not-smoking	Smoking	f1-score
Participant 1	5	0.985	0.995	0.065	0.083
Participant 1	15	0.986	0.995	0.13	0.166
Participant 1	30	0.988	0.996	0.255	0.323
Participant 1	60	0.992	0.998	0.475	0.567
Participant 2	5	0.992	0.997	0.083	0.103
Participant 2	15	0.993	0.998	0.202	0.251
Participant 2	30	0.995	0.998	0.404	0.471
Participant 2	60	0.996	0.999	0.514	0.589
Participant 3	5	0.996	0.999	0.016	0.024
Participant 3	15	0.996	0.999	0.063	0.098
Participant 3	30	0.996	0.999	0.079	0.122
Participant 3	60	0.996	0.999	0.127	0.195
Participant 4	5	0.993	0.997	0	0
Participant 4	15	0.994	0.997	0.106	0.113
Participant 4	30	0.994	0.997	0.136	0.145
Participant 4	60	0.995	0.998	0.318	0.333
Participant 5	5	0.993	0.997	0.06	0.067
Participant 5	15	0.993	0.997	0.072	0.081
Participant 5	30	0.993	0.997	0.145	0.161
Participant 5	60	0.995	0.998	0.301	0.331

5.3 Conclusions

This chapter proposes modelling of smoker's behaviour based on both internal and external factors. The results shows that DL is an effective method to predict smoking events base on an external factor with avrage over all accuracy equals 86.6% , and to some level combining 1D-CNN with CTMoS helps to predict smokers behaviour. The combined model generally achieved good results in predicting the internal craving factors (i.e. nicotine level, nicotine craving, and withdrawal), with average NRMSE 0.20 of predicting of nicotine level in the blood, and 0.24 and 0.23 for both nicotine craving, and withdrawal respectively.

On other hand, work still needs to be done to improve the model performance in term of predicting smoking events base on both internal and external factor. Also, the model is personalised to each participant data. Since the model uses GPS coordinates as one of the external factors it can not be trained to predict other people behaviour. These limitations have been furtherly investigated in the following Chapter 6.

Chapter 6

A 1D-CNN-BiLSTM Model Predicting Smoking Behaviour

This chapter is an improvement on the previously developed smoking behaviour model. it also targets the automatic protection of smoking events base on both internal and external factors.

This chapter presents a smoking behaviour model, which uses the same approach as that investigated in Chapter 5 in that it combines a DL prediction model with CTMoS. However, the model described in this section combines 1D-CNN with BiLSTM to extract feature patterns and long-term dependencies from the external factors. As such, this chapter presents an improved model to the previously developed smoking behaviour model. As in chapter 5, the DL model design will be evaluated first, using the Smartphone Sensors dataset. All 65 days from all 5 participants will be used for validation purposes, and the model will be compared with the SVM, DT, and RF models. Additionally, the model performance will also be compared with 1D-CNN, LSTM, and BiLSTM as separate prediction models. The experimental hypothesis to be tested will be, combining 1D-CNN with BiLSTM gives better performance compared with the other three ML methods and also better than LSTM and BiLSTM and 1D-CNN separately.

The model will be evaluated for its generality and its ability to predict smoking events based on training using data from other smokers, such that each participant's data will be used for testing and the other participants' data will be used for training.

Finally, the model will be evaluated for each participant individually such that for each participant one day will be held for testing, and the other days for training. For all three experiments, k-fold cross-validation will be used for model evaluation.

As in chapter 5, the DL model with the best performance will then be combined with CTMoS. The combined model will be evaluated based on its ability to reflect the internal smoking craving factors (nicotine level, smoking craving, withdrawal). This is in addition to the model ability to predict smoking events in 5- 15- 30- 45- min prior to the appearance of the smoking event.

The DL model uses 4 parameters as input to the model, 3 accelerometer (x, y, and z) data, and coded location input. This new model, termed from here on 1D-CNN-BiLSTM, proves to have a better performance than the other classifiers. And the combined 1D-CNN-BiLSTM - CTMoS can predict smoking behaviour with average f1-score of 0.32, 0.59, 0.71, 0.8 for 5, 15, 30, and 60 -min windows respectively prior to the appearance of the smoking event, which is better than the result achieved by the model described in Chapter 5.

6.1 Data pre-processing

The data that will be used for the model in this chapter is the same Smartphone Sensors dataset (see section 4.5) used earlier in chapter 5. However, in an attempt of improving the model performance, and before using the data, the input vectors were pre-processed as follows.

6.1.1 Coding location data

To make the 14-day pre-quitting data collection as easy as possible, avoid reliance on user input, and avoid influencing the smokers' behaviour by asking them to report many aspects of their behaviour (which is a known behaviour change technique (Michie et al., 2013)), the location data was collected automatically via the smartphone GPS. However, to make the individual coordinates useful for location identification, there is a need to identify locations in a more general way (e.g., an area such as the one around work / home).

Future work should investigate what is the ideal area size that should be used for identification of smoking-locations, but for the purpose of examining whether grouping locations would improve prediction of smoking events, the GPS coordinates were converted to UK postcodes (this could be done for any other coding system based on the country of future users) using an open-source python library (`https://geopy.readthedocs.io/`).

Furthermore, Dunbar et al. (2010), Shiffman et al. (2009, 2014), and Treloar et al. (2014) suggested that there are 5 main sites where a smoker mostly smokes. These are (1) home (2) work or school (3) places when people either eat or socialise like other peoples' homes, bars, restaurants, etc . (4) outdoor locations and (5) others. While the number of locations may differ between individuals, and more work is required to identify the optimal number of locations for each smoker, it was decided to examine whether ranking locations based on how often a smoker smokes in there may improves smoking behaviour prediction.

Based on the data collected for this project, it appeared that identifying the top 5 postcodes where smokers mostly smoke would include the majority of smoking events data. Also, in the Smartphone Sensors dataset in average 71% of the reported smoking locations are in the top 5 smoking locations. It was therefore decided to code the 5 most frequent smoking locations from 1 to 5, where 5 is the most regularly smoked in location. The other locations were grouped with not-smoking locations and coded as 0.

6.1.2 Data portioning

Instead of using input vectors where each sample is the reading for $ith - min$, the one-day samples (1440, one per minute) were portioned into 140 slots, each representing a 10-min window. For example, for the input vector $X_{original} = \{x_1, x_2, x_3, x_4, x_5, ..., x_{1440}\}$, after applying data portioning the vector will be $X_{new} = \{x_{1:10}, x_{11:20}, ..., x_{1431:1440}\}$. The input samples will then become vectors of the accelerometer and location values for each time slot.

If a smoking event occurred during the 10-minute slot (i.e. $x_{i:i+10}$), that specific slot was classified as smoking (1), whereas if not-smoking occurred, it was classified as (0). This approach helped to minimise the overlapping, as well as the imbalance of the labels. The 10-min slot size was selected because it is small enough to be efficient to predict smoking events, but not too big to lose important behaviour patterns.

6.2 Architectural design and evaluation of the 1D-CNNBiLSTM for smoking events prediction

This section describes the general architecture of the 1D-CNN-BiLSTM model of predicting smoking events based on external factors (location and accelerometer). The

proposed model is designed using a multi-headed design approach, in which each sub-model is a combination of BiLSTM and 1D-CNN networks.

While CNN can learn local patterns from input data but it cannot learn sequential correlations; on the other hand, BiLSTM is specialised for sequential modelling, it can extract correlated patterns (see Chapter 3). Therefore, the combination of both networks has the potential to improve the prediction of smoking events, based on input of various external and internal influences on smoking. This has shown good results in several other applications (e.g., rainfall prediction, health condition prediction, etc.) (Liu et al., 2018). Each sub-model in the multi-headed network takes one of the four external factors that can help to characterise (and then predict) the smoking patterns. These factors are: the labelled locations, and the x, y, and z of the accelerometer readings. Each of the input vectors is sampled as a 10-min slot.

In each sub-model, the higher level is 1D-CNN. The CNN consist of an input layer that takes the n-past time slot observations, and an output layer that passes the final feature map to the BiLSTM network. The first hidden layer of the 1D-CNN is the 1D convolutional layer. It uses a sliding convolutional operation over the input vector. The convolutional layer is then followed by a batch normalisation layer, which standardises the input for the next layer. The output of this layer is then passed to the next level of the designed model.

The second level of the sub-model is the BiLSTM network, which consists of several memory units. The BiLSTM layer extracts correlated information from the feature map. When the feature map passes from the 1D-CNN level to the BiLSTM, each of the memory units hidden output is updated, depending on the states of the three gates (input, output, and forgot, as described in section 3.6). The output from the BiLSTM is then passed to a Batch normalisation layer.

The normalised output from each of 4 BiLSTM sub-models is then combined and passed to the final part of the designed model. The final level of the model is made of two layers of fully connected ANN, which are used for generating one hot-encoded output that represents a prediction of either a smoking/not-smoking 10-minute slot. Each of the 2-layers of the fully connected network is followed by a dropout layer. The dropout layer is designed to overcome the over-fitting problem. Figure 6.1 shows the overall architecture of the designed model. ReLU activation function with $l2$ weight regularisation was used for the 1D-CNN, BiLSTM and fully connected layers and *softmax* was used at the output classification layer.

The designed model has several hyperparameters. These include:

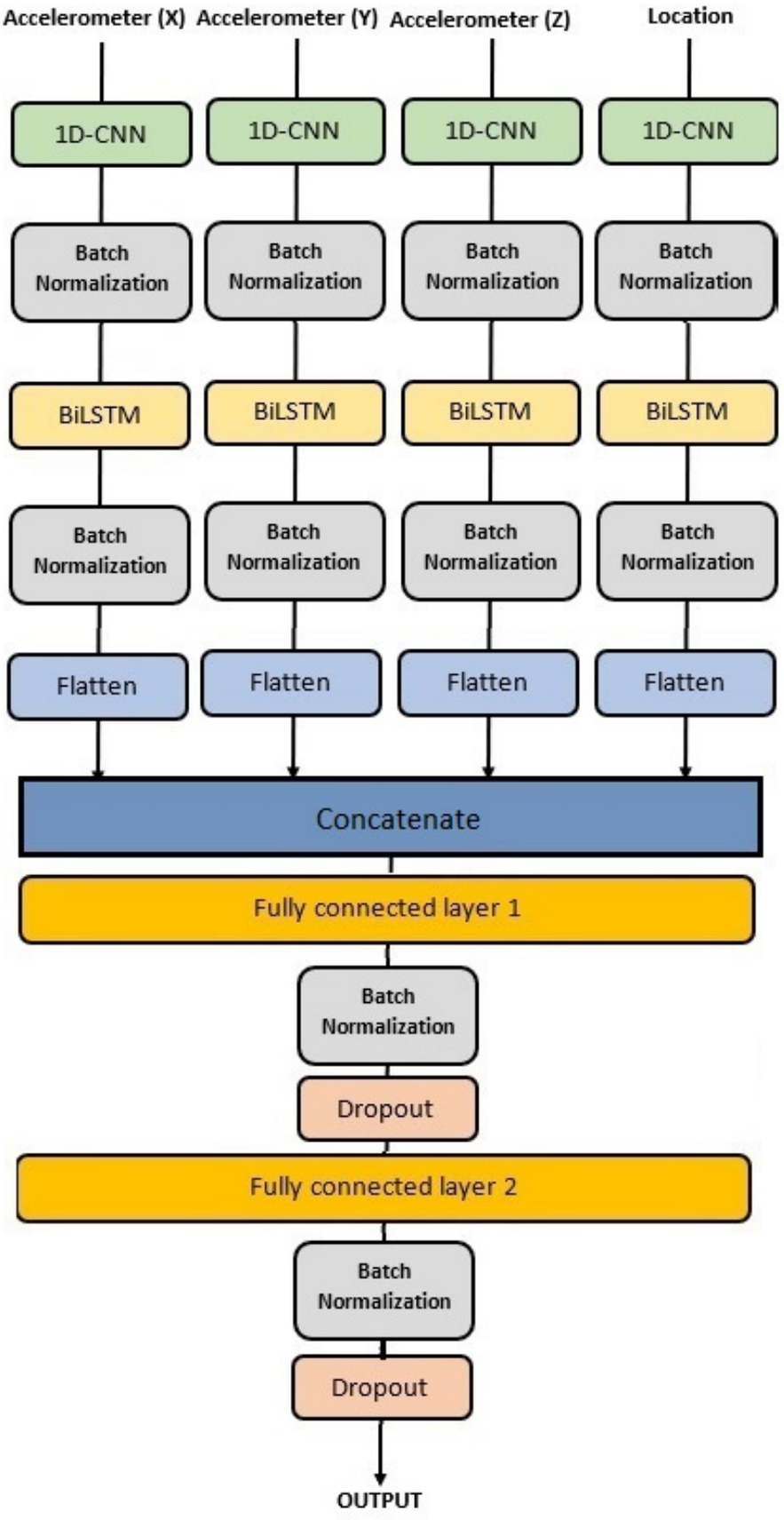


Figure 6.1: The overall architecture of the 1DCNN-BiLSTM smoking events prediction model.

1. n, number of the past observations at the input layer.
2. Two convolutional windows for the 1D-CNN: one for the accelerometer (C1) x, y, and z, and the other for the location (C2).
3. A number of memory units for both accelerometers (M1) values and for location (M2) in BiLSTM network.
4. A number of units for the fully connected ANN in layer one (F1) and layer two (F2).

All of these parameters need to be selected carefully to generate the near-optimum smoking events prediction model. A Genetic Algorithm (GA) is a well-known approach that can be used for hyper-parameter optimization. The GA was implemented using the open-source python DEAP library (https://deap.readthedocs.io/en/master/examples/); this library has been previously used by Sagheer and Kotb (2019) to implement GA that used to select the optimum deep LSTM model to forecast petroleum production.

The GA setting was used similar to that described in Bouktif et al. (2018); which used for hyperparameter selection of LSTM model of electric load forecasting the selected population size and the number of generations were set to 40.

- Crossover: two points of crossover with crossover probability equals to 0.7.
- Mutation: is flip with a probability of 0.1.

Validation accuracy was used as the fitness function for the GA, in which individuals with the highest accuracy were selected. The search range for (n) between 2 to 16, and for the C1, C2, M1, M2, F1, and F2 was in the range between 16 to 512. Table 6.1 describes the final model with the final selected value for each of the optimised hyperparameters.

Table 6.1: Detailed description of 1D-CNN-BiLSTM architecture.

Layer (type)	**Output Shape**	**Param #**	**Connected to**
input1	[(None, 11, 10)]	0	
input2	[(None, 11, 10)]	0	
input3	[(None, 11, 10)]	0	
input4	[(None, 11, 10)]	0	
Conv1 (Conv1D)	(None, 10, 73)	1533	input1
Conv2 (Conv1D)	(None, 10, 73)	1533	input2
Conv3 (Conv1D)	(None, 10, 73)	1533	input3
Conv4 (Conv1D)	(None, 10, 60)	1260	input4
batch normalizationL11	(None, 10, 73)	292	Conv1 (Conv1D)
batch normalizationL12	(None, 10, 73)	292	Conv2 (Conv1D)
batch normalizationL13	(None, 10, 73)	292	Conv3 (Conv1D)
batch normalizationL14	(None, 10, 60)	240	Conv4 (Conv1D)
bidirectional1	(None, 10, 200)	139200	batch normalizationL11
bidirectional2	(None, 10, 200)	139200	batch normalizationL12
bidirectional3	(None, 10, 200)	139200	batch normalizationL13
bidirectional4	(None, 10, 308)	264880	batch normalizationL14
batch normalizationL21	(None, 10, 200)	800	bidirectional1
batch normalizationL22	(None, 10, 200)	800	bidirectional2
batch normalizationL23	(None, 10, 200)	800	bidirectional3
batch normalizationL24	(None, 10, 308)	1232	bidirectional4
Flatten1	(None, 2000)	0	batch normalizationL21
Flatten2	(None, 2000)	0	batch normalizationL22
Flatten3	(None, 2000)	0	batch normalizationL23
Flatten4	(None, 3080)	0	batch normalizationL24
concatenate (Concatenate)	(None, 9080)	0	Flatten1 Flatten2 Flatten3 Flatten4
Dense layer1	(None, 156)	1416636	concatenate
batch normalization D1	(None, 156)	624	Dense layer1
Dropout D1	(None, 156)	0	batch normalization D1
Dense layer2	(None, 112)	17584	Dropout D1
batch normalization D2	(None, 112)	448	Dense layer2
Dropout D2	(None, 112)	0	batch normalization D2
output (Dense)	(None, 2)	226	Dropout D2

6.2.1 Model evaluation

The Smartphone Sensors dataset (described in Chapter 4) was used for the model evaluation, using k-folds cross-validation. EarlyStopping and ModelCheckpoint (python Keras functions) were used to improve the training process. The classification model was tested to see whether the classifier could accurately detect the daily 10-min smoking slots based on the smoker's motion and location. The input vector, as explained in the previous section, was the 3 accelerometer values and coded location values, and the output is a prediction of (0) for not-smoking and (1) for smoking for each of the 10-min slots.

6.2.2 Results for the general-model performance

In order to test the efficiency of the designed model, the model was compared with SVM, DT, RF, 1D-CNN, LSTM, and BiLSTM. First, the model is evaluated with 64 days, using the k-folds validation process. In each fold, one day is used for testing and the other 63 days are used for training. Each day consist of 140 samples of 10-minutes windows. For SVM, DT, and RF, the same optimization method and setting as those described in chapter 5 were used here to validate the models. GA was applied for 1D-CNN, LSTM, BiLSTM, and 1D-CNN-BiLSTM for hyperparameters optimization.

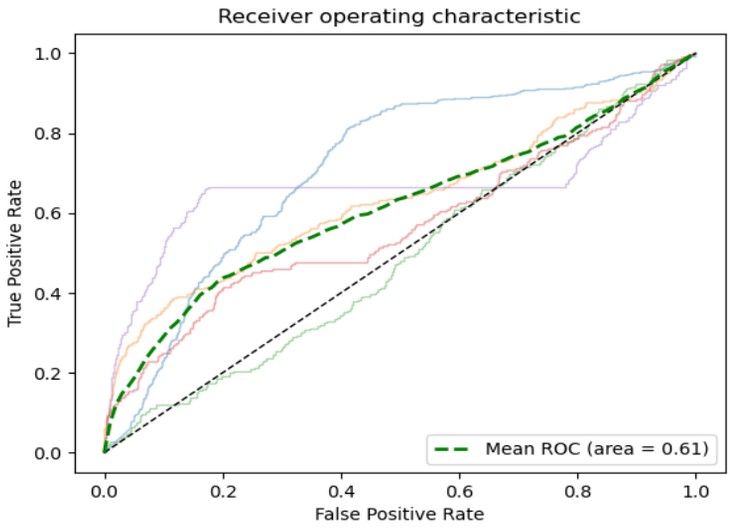
Model accuracy (equation 3.1), and f1-score (equation 3.4) are all calculated to evaluate the model performance. Table 6.2 show the evaluation metrics of the 1D-CNN-BiLSTM model, along with all the other models. The models were evaluated for each day, and the average of all 64 folds is calculated and presented in the table.

Table 6.2: Models evaluation metrics for smoking behaviour prediction based on location and motion input.

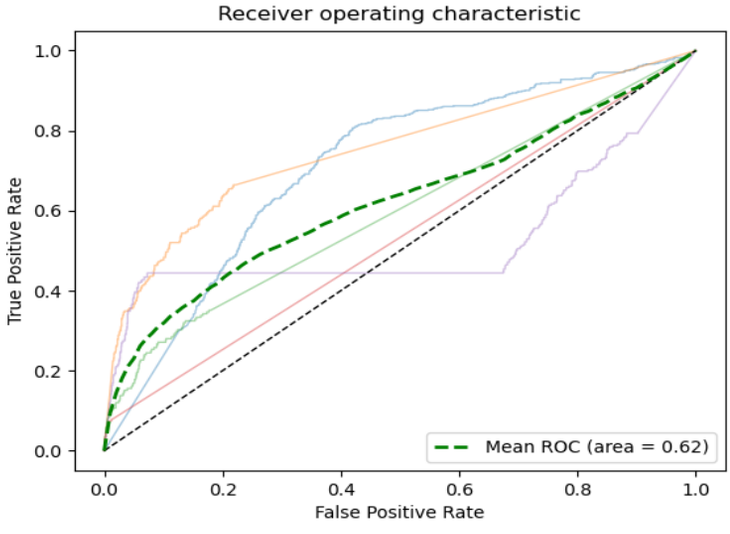
Model type	Accuracy	Not-smoking	Smoking	f1-score
SVM	0.813	**0.993**	0.008	0.015
DT	0.563	0.559	0.55	0.317
RF	0.628	0.636	**0.593**	0.368
CNN	0.863	0.953	0.425	0.49
LSTM	0.866	0.949	0.462	0.527
BiLSTM	0.869	0.946	0.487	0.546
1D-CNN-BiLSTM	**0.872**	0.951	0.477	**0.548**

To test the hypothesis of CNNBiLSTM is better in performance than the other prediction models, as in Chapter 5, the Friedman test was used as statistical significant test methods. The test is be applied on the models' overall accuracy, comparing architectures (CNNBiLSTM compared against SVM, DT, RF, CNN, LSTM, BiLSTM). This yeilds $Q = 18.6, d.f. = 4, p = 0.00198$ which is less than both α values of 0.05 and 0.1. Thus the models are significantly different in their performance. Comparing models, generally both BiLSTM and 1D-CNN-BiLSTM were better in predicting smokers' behaviour based on external factors.

In order to test the model generality in predicting other smokers' behaviour, the BiLSTM and 1D-CNN-BiLSTM was evaluated for each participant individually, using the k-fold validation process. In each model, one participant's data was used for testing, and the other 4 participants' data was used for training. Figure 6.2 shows the Receiver Operating Characteristic (ROC) performance for the two models.



(a) ROC curve for BiLSTM.



(b) ROC curve for 1D-CNNBiLSTM.

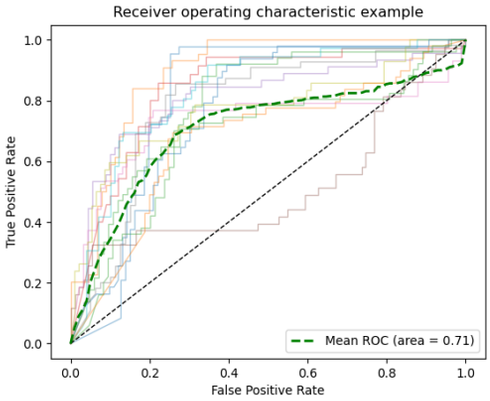
Figure 6.2: The ROC curve. The shaded lines are the ROC curve for each participant separately and the dashed green line is the mean ROC curve.

It can be seen from all the results that 1D-CNN-BiLSTM has higher average accuracy (87.2%) and f1-score (54.8%). In addition, when trying to predict smoking events and training the model on other smokers data, the combination of BiLSTM with 1D-CNN made the 1D-CNN-BiLSTM model better in predicting the behaviour of other smokers, with mean Area Under the Curve (AUC) of 0.62. This is an important advantage if this is to be used for an app to the public, as it means that predictions can be

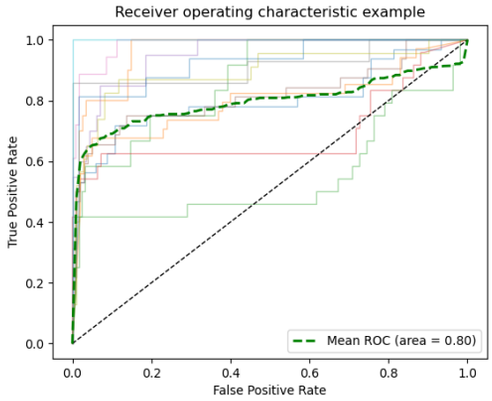
improved for individuals, by using a larger dataset from other users.

6.2.3 Results for the personalised-model performance

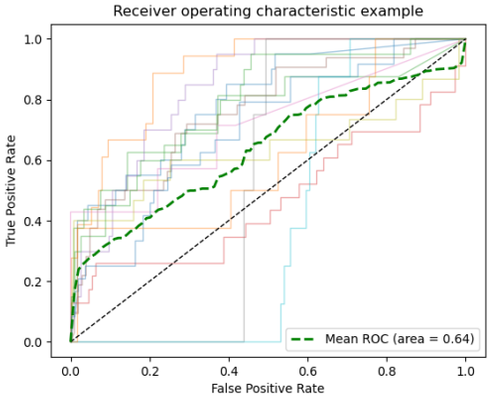
The 1D-CNN-BiLSTM proved to provide a better prediction than the other models. The 1D-CNN-BiLSTM will be tested to identify its performance in predicting smoking behaviour for each participant individually. To this end, for each participant one day was used for testing, and the other 11 - 12 days (one of the participants has only 12 days instead of 13) were used for training. This approach can show the model ability in predicting smoking events based on one person's data. Figure 6.3 shows the ROC for each participant.



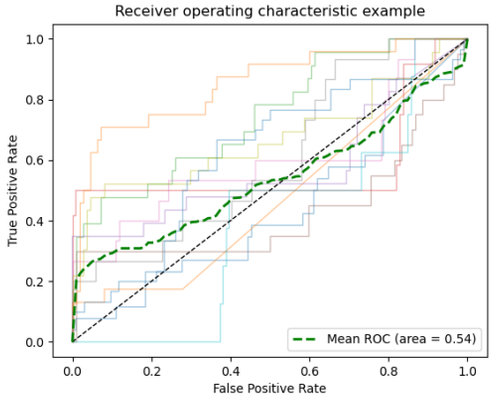
(a) participant 1.



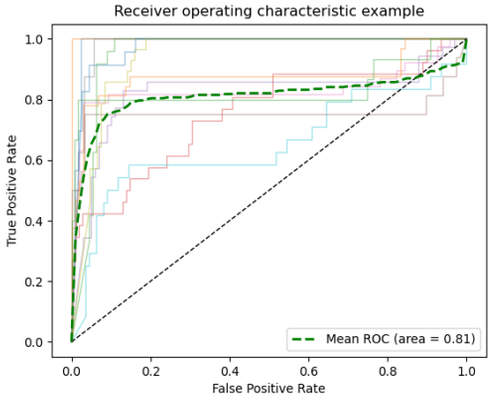
(b) participant 2.



(c) participant 3.



(d) participant 4.



(e) participant 5.

Figure 6.3: The ROC curve for the 1D-CNN-BiLSTM trained for each participant individually. The shaded lines are the ROC curve for each individual day; the green dashed line is the average of all days

While the overall performance of the general-model is good, It seems that is very much affected by the engagements of participants: performance is much better when the participants are engaging well with the app, and when the participants' data has a low level of app-off samples.

On another hand, it seems that the model is better at predicting the smoking events if it is trained to each participant individually (overall accuracy 88.1% and 57.3% accuracy of predicting smoking events). This result is expected due to the individual differences in peoples' behaviour.

To conclude, BiLSTM is better for predicting smoking events, and combining it with 1D-CNN can improve the overall model performance. Smokers' motion and location are important factor that can help to understand the smoker's behaviour (see Chapter 5), and they are easily collected using smartphone sensors. Moreover, the performance of the 1D-CNN-BiLSTM is better than the 1D-CNN (overall average accuracy of 86.6%, and 38.6% predicting smoking events) described in Chapter 5.

Finally, coding GPS data can help to generalise the model, but more work may be needed to investigate different possibilities of coding (i.e., could there be better classifications than postcodes, and whether 5 locations are optimal). To have a better understanding of the performance of the general model, it needs to be tested on a larger dataset, which is the target for future work on this project.

6.3 Using 1D-CNN-BiLSTM model to predict a smoker's behaviour

Internal factors that influence smoker behaviour (i.e., nicotine level, craving and withdrawal symptoms level) can be modelled using CTMoS. The same approach of combining DL (now BiLSTM-CNN) with CTMoS is also going to be used in this chapter. Figure 6.4 shows the architecture of 1D-CNN-BiLSTM combined with CTMoS.

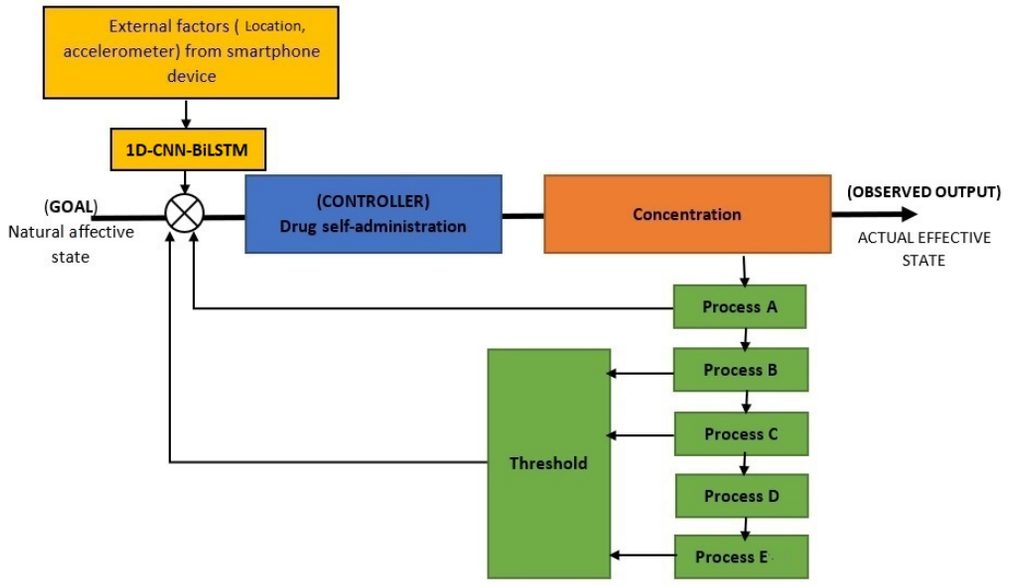
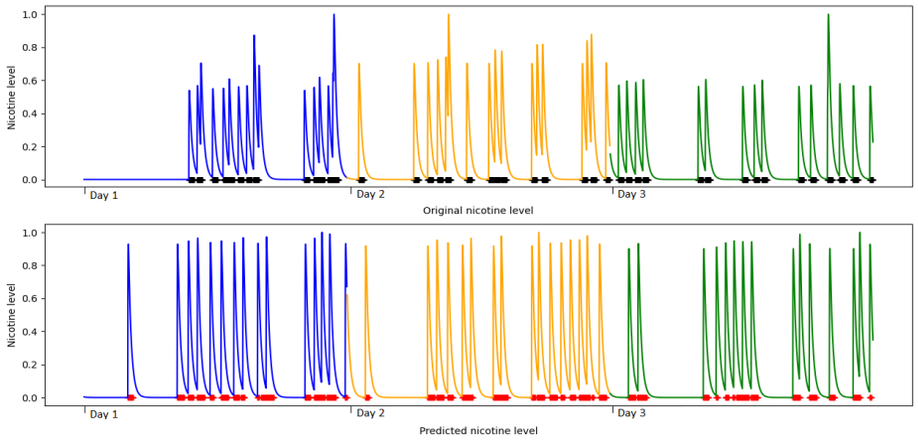


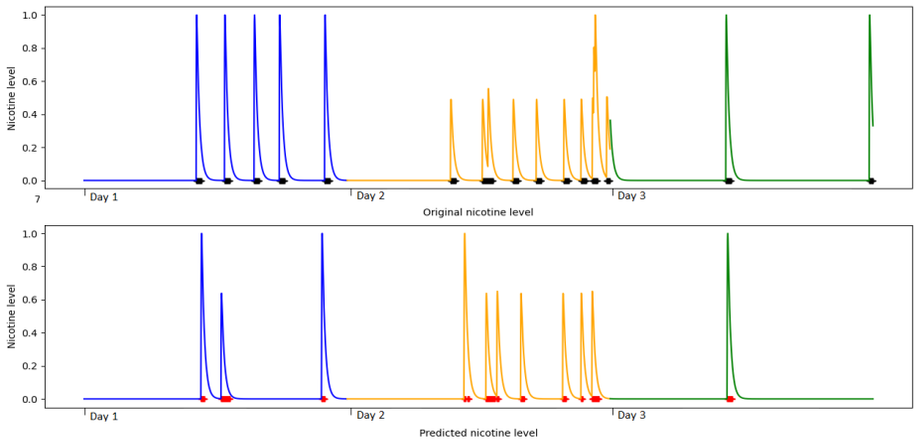
Figure 6.4: Smoking behaviour model 1D-CNN-BiLSTM. 2 predictors are used as input to the 1D-CNN-BiLSTM model. A classification value of 1 represents a potential smoking event. This value passed to the CONTROLLER, simulating the taking of a cigarette, and re-initializing the parameters of the control model to zero.

The 1D-CNN-BiLSTM makes use of accelerometer and coded location values to predict the likelihood of a smoking event. Since the CTMoS calculates internal factors that influence behaviour for every minute, and 1D-CNN-BiLSTM predicts 10-min smoking slot, the one value of the 1D-CNN-BiLSTM is extended to fill all the 10-mins. Whenever a 10-minute slot is considered as the smoking period, the CTMoS checks if the nicotine level is less than the threshold, which can be an indicator of a potential smoking event. The model was tested for each participant separately.

Figure 6.5, Figure 6.7, and Figure 6.6 show randomly selected days for two participants, predicting nicotine, craving and withdrawal symptoms levels, respectively. One participant has a good level of engagement and reporting of smoking events, while the second participant has a low level of reported smoking events. The figures show a comparison between the CTMoS curves calculated using the reported smoking events from the Smartphone Sensors dataset, and the predicted values of nicotine, craving and withdrawal symptoms levels using the 1D-CNN-BiLSTM model.

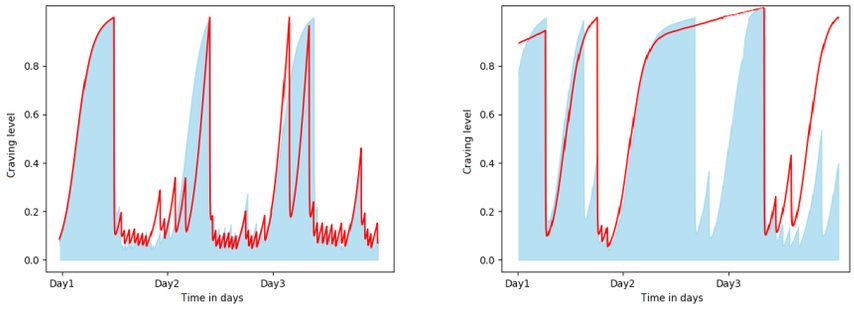


(a) Example 1: of randomly selected three day sequence from one participant.



(b) Example 2: of randomly selected three day sequence from one participant that has days with high 'app-off' data.

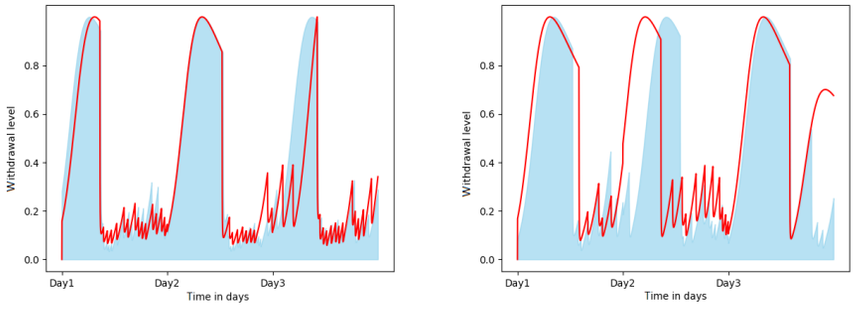
Figure 6.5: Results of the hybrid smoking behaviour model; the output is for three randomly selected days from two different participants, each different colour present different day. the black + markers are the real reported smoking events from the original data, and the red + markers are the predicted smoking events 1D-CNN-BiLSTM



(a) Example 1: three consecutive days from one participant.

(b) Example 2: three consecutive days from one participant that has days with with high 'app-off' data.

Figure 6.6: A comparison between craving level (the blue shaded area) generated using reported smoking events, and craving level predicted using the designed model (the red line).



(a) Example 1: three consecutive days from one participant.

(b) Example 2: three consecutive days from one participant that has days with with high 'app-off' data.

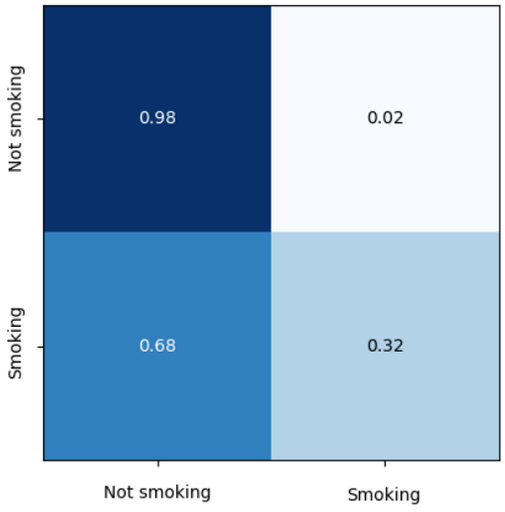
Figure 6.7: A comparison between withdrawal level (the blue shaded area) generated using reported smoking events, and withdrawal level predicted using the designed model (the red line).

In general, the model is reliable in predicting the internal factors that influence smoking behaviour; MSE and RMSE were calculated to understand the error rate in predicting internal behaviour as shown in Table 6.3.

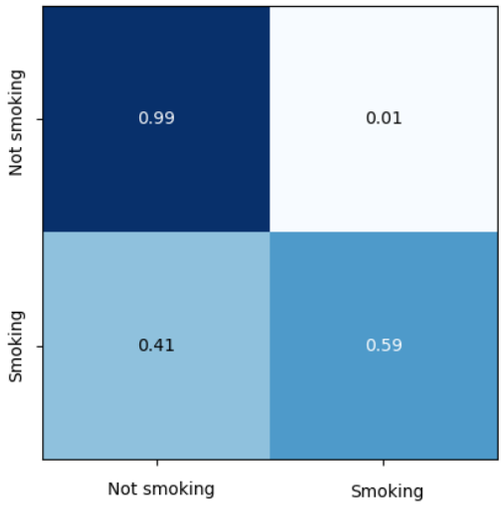
Table 6.3: The overall error rate of the designed model of predicting nicotine level.

	Nicotine level		Craving level		Withdrawal level	
	MSE	RMSE	MSE	RMSE	MSE	RMSE
Participant 1	0.071	0.261	0.022	0.138	0.032	0.165
Participant 2	0.03	0.168	0.056	0.228	0.064	0.237
Participant 3	0.025	0.16	0.092	0.285	0.076	0.256
Participant 4	0.03	0.173	0.164	0.396	0.138	0.349
Participant 5	0.043	0.203	0.118	0.336	0.105	0.312

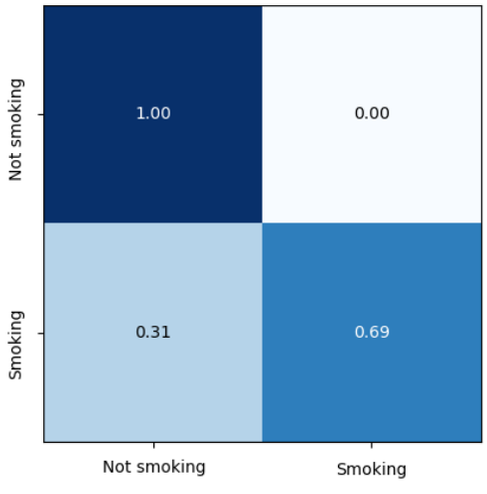
The following Figure 6.8 shows the confusion matrix for the combined 1D-CNN-BiLSTM and CTMoS. The model predicts smoking events based on external factor using 1D-CNN-BiLSTM and internal factors using CTMoS in 5 min, 15 min, 30 min, and 1-h window prior to the appearance of the smoking events. Table 6.4 shows the accuracy of the designed model in a 5 min, 15 min, 30 min, and 1-h window prior to the appearance of the smoking events; along with f1-score for the same time windows.



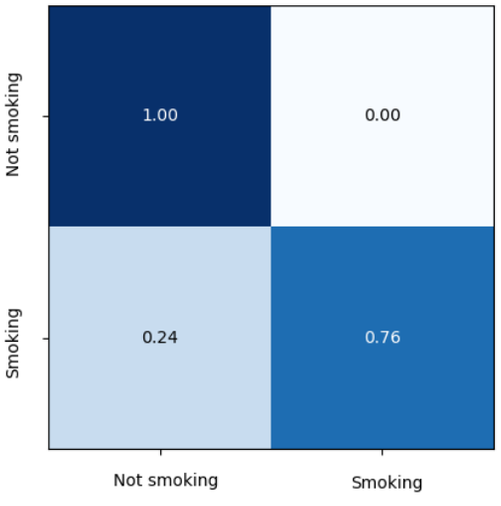
(a) 5 min prior to the smoking event.



(b) 15 min prior to the smoking event.



(c) 30 min prior to the smoking event.



(d) 1-h min prior to the smoking event.

Figure 6.8: Overall confusion matrix of predicting smoking events using the combined model in 5 min, 15 min, 30 min, and 1-h window prior to the appearance of the smoking events

Table 6.4: The evaluation matrix for all participants, shows the prediction accuracy of the combined model.

PID	time window	accuracy	Not-smoking	Smoking	f1-score
Participant 1	5	0.926	0.966	0.412	0.442
Participant 1	15	0.97	0.989	0.716	0.771
Participant 1	30	0.983	0.995	0.836	0.877
Participant 1	60	0.988	0.997	0.88	0.915
Participant 2	5	0.958	0.986	0.283	0.35
Participant 2	15	0.981	0.997	0.608	0.72
Participant 2	30	0.988	0.999	0.729	0.828
Participant 2	60	0.991	1	0.781	0.872
Participant 3	5	0.972	0.993	0.216	0.29
Participant 3	15	0.98	0.996	0.386	0.505
Participant 3	30	0.986	0.999	0.506	0.653
Participant 3	60	0.99	0.999	0.639	0.767
Participant 4	5	0.971	0.993	0.115	0.164
Participant 4	15	0.977	0.996	0.228	0.33
Participant 4	30	0.98	0.997	0.304	0.423
Participant 4	60	0.986	0.999	0.438	0.598
Participant 5	5	0.969	0.981	0.42	0.358
Participant 5	15	0.982	0.987	0.761	0.639
Participant 5	30	0.989	0.993	0.814	0.751
Participant 5	60	0.993	0.996	0.845	0.83

6.4 Conclusions

In conclusion, the designed model can reliably predict smoking events based on both internal and external factors; it can be reliably used in the future in the development of smoking cessation apps, where it will entirely relay only on data that collected from smartphone sensors to predict smoker's behaviour.

To date, there have been only a few studies that used ML in the automatic prediction of smoking behaviour (e.g. Schick et al. (2018)). However, no published study to date had designed a ML model to combine the internal and external factors that influence smokers' behaviour. Generally, most of the previous ML models rely on self-reporting

of predictors (e.g. craving for smoking, location of smoking, being near other smokers, etc.) in the quitting period (Dumortier et al., 2016; Koslovsky et al., 2018b), in order to predict potential relapse risk. This approach has proved to be inefficient, due to lack of reliability of self-reporting (Businelle et al., 2016; Webb et al., 2020), and also the need to report increasing number of factors, which makes it unfeasible for regular use (see Chapter 2 for review).

One piece of research that uses automatic prediction of smoking events using ML without self-reporting, which is much closer the designed model is MapMySmoke (Schick et al., 2018), However, this study did not report any analytical result regarding the prediction model, only positive feedback from the participants.

Finally, although the overall performance of the model seems good, a larger dataset needs to be explored in the future to have a better understanding of the model performance and limitations.

Chapter 7

Conclusions and Future work

7.1 Conclusions

The main objective of this research was to design a deep learning model that can automatically learn smokers' behaviour with minimal user input, and use this to predict smoking events. The hope is that this (or a similar) model could be used in the future in the development of smart smoking cessation app.

A literature review (Chapter 2, objective number 1, contribution 1) had identified and critically evaluated studies that are related to the scope of this project. Based on this review, the main approaches for designing smoking cessation apps, and their limitations have been identified, including apps that use the EMA method to collect external craving factors and sent target interventions (Businelle et al., 2016; Hébert et al., 2018; Naughton et al., 2016). The review also described the importance of internal factors on smoking behaviour (Stevenson et al., 2017) and how it can be modelled using opponent-process theory (Gutkin and Ahmed, 2011). Also, the work by Bobashev et al. (2017) was highlighted due to its ability to mathematically to model the smoking behaviour and the internal factors (i.e. nicotine level, craving and withdrawal) without relying on any complex any neurobiological process.

Based on the literature it has been concluded that both internal and external factors are involved in governing the smoking patterns and both factors should be included when trying to develop smoking cessation apps. Combining these methods with the psychological literature, the review also noted that the reliability of smoking cessation apps can be improved by minimising the reliance on self-reporting, which can be done by full use of advanced technology as recently provided by smartphone devices.

Following the review, it was decided to combine CTMoS with deep learning methods, in order to enable the combined modelling of internal and external factors that influence smokers' behaviour.

First, it was important to validate the CTMoS model (Chapter 4, objective number 2, contribution 2), as it was previously only validated using animal data (Bobashev et al., 2017). The QSense dataset (Naughton et al., 2016), was identified as appropriate for validation of CTMoS, as it included craving level as reported by smokers using EMA. Results of this work had shown that CTMoS can reliably reflect the overall increase and decrease in smokers' cue-induced craving during the day, especially when the smoker experiences low craving for smoking. However, as human smoking behaviour is also influenced by external cues (Grainge et al., 2009; Hughes et al., 2017; Shiffman et al., 2014), CTMoS is not sufficient for accurately modelling human smoking behaviour.

To capture the influence of external factors, a second data set was collected (chapter 4, objective number 3). This data set included smartphone sensors dataset, and was collected from 5 participants over a 14-day period, for this research. The GPS and accelerometer data in the dataset were used to model the external influences on smokers' behaviour using deep learning algorithms.

Two deep learning models have been tested, and combined with CTMoS in order to examine their effectiveness for predicting smoking behaviour (chapter 5 and 6, objective number 4, contribution 3). The resultant models used CTMoS to model nicotine level, and related craving, and withdrawal symptoms levels, and combined this with external factors (i.e., accelerometer and location data) collected using smokers' personal smartphone device. The two designed model were implemented using Keras API, which is a python open-source high-level API for TensorFlow (`https://www.tensorflow.org/api_docs/python/tf/keras`) For SVM, DT, RF and the evaluation matrixes the free python machine learning library scikit-learn was used (`https://scikit-learn.org/stable/index.html`).

The first deep learning model for predicting smoking events was 1D-CNN. The designed model uses raw accelerometer and GPS data. The model was tested for each participant individually. The designed model combined 1D-CNN with CTMoS to predict smoking behaviour with average f1-score equal to 0.06, 0.14, 0.24, 0.4 for 5, 15, 30, and 60 -min respectively, prior to the appearance of the smoking event based on both the internal nicotine level and external factors. The model achieved good results in predicting the internal craving factors (i.e. nicotine level, nicotine craving, and

withdrawal); but it's smoking events prediction needs to be improved.

The second model has improved the performance of the first model by combining BiLSTM with 1D-CNN. Combining 1D-CNN with BiLSTM enabled the extraction of higher feature patterns using 1D-CNN, while also extracting the sequential correlations in the input sequences with BiLSTM. The model's hyper-parameters were selected using Genetic Algorithm to achieve a near-optimal model. In order to generalise the designed model, the location data was replaced by coded data that reflect ranked 5 locations where the user is most likely to smoke in. However, future work should examine whether better classification for location data (i.e., radios of e.g., 100 meters instead of postcodes) could improve performance, and whether an adjustable number of top locations can be used to individualise the algorithm to different smokers' habits of smoking (e.g., those who have many locations they smoke in, compared to those who have very few). Combining the 1D-CNN-BiLSTM with CTMoS enabled the prediction of smoking events to be base on both internal and external factors, the model was able to predict smoking behaviour with average f1-score of 0.32, 0.59, 0.71, 0.8 for 5, 15, 30, and 60 -min windows respectively, prior to the appearance of the smoking event.

Overall, work in this thesis represents significant advances to the modelling of smoking behaviour in order to assist smokers in quitting. All previous reported studies in literature, even if they used ML models, they have relied on self-reporting in post-quitting period (Dumortier et al., 2016; Koslovsky et al., 2018b). The other work that did not use self-reporting (Schick et al., 2018) they used different experiment settings and different inputs and their reported analysis is not sufficient to compare with the designed model.

This model can be used to develop a smoking cessation app. Using the advanced DL method enables reliably sending automated, timely and targeted intervention messages to smokers who wish to quit smoking.

7.2 Findings and the impact of the project

This research aims to develop an automatic smoking event prediction model based on both internal and external factors.

It has been found from previous works that both internal and external factors influence smoking behaviour and smoking craving. The literature review also indicated

that there are two main limitations in the used approaches in designing smoking cessation apps; first using self-reporting of craving factors and also, it lacks on its ability to understanding both internal and external factors. the literature review also highlighted the importance of employing ML as a method in predicting smoking events.

After validating the CTMoS using human participant data, It has been found the CTMoS can be a reliability used in modelling internal factors especially when the aim is to reflect the general increase and decrease in craving level. It is also expected that the CTMoS may fail in accurately predicting smoking craving due to some external factors.

It is also found that the DL approach can be accurately used in protecting smoking events based on external factors. And combing DL with the CTMoS will help to capture both internal and external factors that influence the smoker's behaviour.

The designed model opens the door to the possibility of automatic prediction the smoking and other problematic behaviour, which will in turn enable sending automated, timely and targeted intervention messages to smokers who wish to quit smoking, based on their individuals' behaviour. Future work can also include model of understanding the smokers' motivation, or other barriers, and personalise messages even further.

7.3 Future work

While this work shows promising results in modelling the smoker's behaviour, there is an important limitation of the model suggested. Although the designed model was able to predict smoking events while the user is still smoking, there remains a lack of understanding of craving behaviour and potential lapses when the smoker decides to quit. In other words, more work is required to understand what parameters need to be collected and how they should be modelled.

Another improvement in the designed model is to collect additional data, like WiFi signal, detection of Bluetooth devices; also testing other external factors like the type of activity (e.g. sitting, walking, driving, etc.), and indoor/outdoor prediction.

The model can also be improved to be a general model (i.e. generalising external factors), more specifically coding the collected GPS coordinates, either by asking the participant to label their smoking location such as work and home, as well as public locations such as bars and tobacco shops, which are likely to be associated with smoking or do it automatically by using special Android libraries that enables categorise

the GPS data into (e.g. bars, restaurants, private places, etc.).

Moreover, The designed model is based on modelling smoker's behaviour inside the UK. Future work is to test the suggested model on a larger population with different cultural background.

References

Abdoli, Sajjad, Patrick Cardinal, and Alessandro Lameiras Koerich (2019). "End-to-end environmental sound classification using a 1D convolutional neural network". In: *Expert Systems with Applications* 136, pp. 252–263.

Abo-Tabik, Maryam, Yael Benn, and Nicholas Costen (2021). "Are Machine Learning Methods the Future for Smoking Cessation Apps?" In: *Sensors* 21.13. ISSN: 1424-8220. DOI: 10.3390/s21134254.

Abo-Tabik, Maryam, Nicholas Costen, John Darby, and Yael Benn (2019). "Decision Tree Model of Smoking Behaviour". In: *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, pp. 1746–1753.

— (2020). "Towards a smart smoking cessation app: a 1D-CNN model predicting smoking events". In: *Sensors* 20.4, p. 1099.

Abroms, Lorien C, J Lee Westmaas, Jeuneviette Bontemps-Jones, Rathna Ramani, and Jenelle Mellerson (2013). "A content analysis of popular smartphone apps for smoking cessation". In: *American journal of preventive medicine* 45.6, pp. 732–736.

Agarap, Abien Fred (2018). "Deep learning using rectified linear units (relu)". In: *arXiv preprint arXiv:1803.08375*.

Ahmed, Serge H, Georgiy Bobashev, and Boris S Gutkin (2007). "The simulation of addiction: pharmacological and neurocomputational models of drug self-administration." In: *Drug and alcohol dependence* 90.2-3, pp. 304–311.

Babb, Stephen, Ann Malarcher, Gillian Schauer, Kat Asman, and Ahmed Jamal (2017). "Quitting smoking among adults—United States, 2000–2015". In: *Morbidity and Mortality Weekly Report* 65.52, pp. 1457–1464.

Barros, Rodrigo Coelho, Márcio Porto Basgalupp, Andre CPLF De Carvalho, and Alex A Freitas (2011). "A survey of evolutionary algorithms for decision-tree induction". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.3, pp. 291–312.

Baskerville, Neill Bruce, Laura Louise Struik, David Hammond, G Emmanuel Guindon, Cameron D Norman, Robyn Whittaker, Catherine M Burns, Kelly A Grindrod, and K Stephen Brown (2015). "Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial study protocol". In: *JMIR Research Protocols* 4.1, e10.

Belgiu, Mariana and Lucian Drăguţ (2016). "Random forest in remote sensing: A review of applications and future directions". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 114, pp. 24–31.

Bobashev, Georgiy, John Holloway, Eric Solano, and Boris Gutkin (2017). "A Control Theory Model of Smoking". In: *Methods report (RTI Press)* 2017.

Bonissone, Piero, José M Cadenas, M Carmen Garrido, and R Andrés Dıaz-Valladares (2010). "A fuzzy random forest". In: *International Journal of Approximate Reasoning* 51.7, pp. 729–747.

Bouktif, Salah, Ali Fiaz, Ali Ouni, and Mohamed Adel Serhani (2018). "Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches". In: *Energies* 11.7, p. 1636.

— (2020). "Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting". In: *Energies* 13.2, p. 391.

Breiman, L, JH Friedman, RA Olshen, et al. (2017). "Classification and regression trees Routledge". In:

Breiman, Leo (1996). "Bagging predictors". In: *Machine learning* 24.2, pp. 123–140.

Breiman, Leo, Jerome H Friedman, Richard A Olshen, and Charles J Stone (1984). "Classification and regression trees. Belmont, CA: Wadsworth". In: *International Group* 432, pp. 151–166.

Briem, Gunnar Jakob, Jon Atli Benediktsson, and Johannes R Sveinsson (2002). "Multiple classifiers applied to multisource remote sensing data". In: *IEEE transactions on geoscience and remote sensing* 40.10, pp. 2291–2299.

Buller, David B, Ron Borland, Erwin P Bettinghaus, James H Shane, and Donald E Zimmerman (2014). "Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation". In: *Telemedicine and e-Health* 20.3, pp. 206–214.

Businelle, Michael S, Ping Ma, Darla E Kendzor, Summer G Frank, David W Wetter, and Damon J Vidrine (2016). "Using intensive longitudinal data collected via mobile phone to detect imminent lapse in smokers undergoing a scheduled quit attempt". In: *Journal of medical Internet research* 18.10, e275.

Cambron, Christopher, Aaron K Haslam, Brian RW Baucom, Cho Lam, Christine Vinci, Paul Cinciripini, Liang Li, and David W Wetter (2019). "Momentary precipitants connecting stress and smoking lapse during a quit attempt." In: *Health Psychology*.

Camps, Julia, Albert Sama, Mario Martin, Daniel Rodriguez-Martin, Carlos Perez-Lopez, Joan M Moreno Arostegui, Joan Cabestany, Andreu Catala, Sheila Alcaine, Berta Mestre, et al. (2018). "Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit". In: *Knowledge-Based Systems* 139, pp. 119–131.

Cao, Jian, Zhi Li, and Jian Li (2019). "Financial time series forecasting model based on CEEMDAN and LSTM". In: *Physica A: Statistical Mechanics and its Applications* 519, pp. 127–139.

Chakravarthi, Bharathi Raja, Ruba Priyadharshini, Vigneshwaran Muralidaran, Shardul Suryawanshi, Navya Jose, Elizabeth Sherly, and John P McCrae (2020). "Overview of the track on sentiment analysis for dravidian languages in code-mixed text". In: *Forum for Information Retrieval Evaluation*, pp. 21–24.

Chandra, Siddharth, Deborah Scharf, and Saul Shiffman (2011). "Within-day temporal patterns of smoking, withdrawal symptoms, and craving". In: *Drug and alcohol dependence* 117.2-3, pp. 118–125.

Chandra, Siddharth, Saul Shiffman, Deborah M Scharf, Qianyu Dang, and William G Shadel (2007). "Daily smoking patterns, their determinants, and implications for quitting." In: *Experimental and clinical psychopharmacology* 15.1, p. 67.

Chang, Yin-Wen, Cho-Jui Hsieh, Kai-Wei Chang, Michael Ringgaard, and Chih-Jen Lin (2010). "Training and testing low-degree polynomial data mappings via linear SVM." In: *Journal of Machine Learning Research* 11.4.

Chen, Fanglin, Rui Wang, Xia Zhou, and Andrew T Campbell (2014). "My smartphone knows i am hungry". In: *Proceedings of the 2014 workshop on physical analytics*, pp. 9–14.

Chen, Hamid Reza Pourghasemi, Aiding Kornejady, and Ning Zhang (2017). "Landslide spatial modeling: Introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques". In: *Geoderma* 305, pp. 314–327.

Creamer, MeLisa R, Teresa W Wang, Stephen Babb, Karen A Cullen, Hannah Day, Gordon Willis, Ahmed Jamal, and Linda Neff (2019). "Tobacco product use and cessation indicators among adults—United States, 2018". In: *Morbidity and mortality weekly report* 68.45, p. 1013.

De Mántaras, R López (1991). "A distance-based attribute selection measure for decision tree induction". In: *Machine learning* 6.1, pp. 81–92.

Deka, Bhabesh, Pradipta Maji, Sushmita Mitra, Dhruba Kumar Bhattacharyya, Prabin Kumar Bora, and Sankar Kumar Pal (2019). *Pattern Recognition and Machine Intelligence: 8th International Conference, PReMI 2019, Tezpur, India, December 17-20, 2019, Proceedings, Part I*. Vol. 11941. Springer Nature.

Demšar, Janez (2006). "Statistical comparisons of classifiers over multiple data sets". In: *The Journal of Machine Learning Research* 7, pp. 1–30.

Dumortier, Antoine, Ellen Beckjord, Saul Shiffman, and Ervin Sejdić (2016). "Classifying smoking urges via machine learning". In: *Computer methods and programs in biomedicine* 137, pp. 203–213.

Dunbar, Michael S, Deborah Scharf, Thomas Kirchner, and Saul Shiffman (2010). "Do smokers crave cigarettes in some smoking situations more than others? Situational correlates of craving when smoking". In: *Nicotine & Tobacco Research* 12.3, pp. 226–234.

Enquist, Magnus and Stefano Ghirlanda (2005). *Neural networks and animal behavior*. Vol. 33. Princeton University Press.

Erickson, Bradley J, Panagiotis Korfiatis, Zeynettin Akkus, and Timothy L Kline (2017). "Machine learning for medical imaging". In: *Radiographics* 37.2, pp. 505–515.

Ferdous, Raihana, Venet Osmani, and Oscar Mayora (2015). "Smartphone app usage as a predictor of perceived stress levels at workplace". In: *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, pp. 225–228.

Fibla, Marti Sanchez, Ulysses Bernardet, and Paul FMJ Verschure (2010). "Allostatic control for robot behaviour regulation: An extension to path planning". In:

2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 1935–1942.

Fraser, Alex S (1957). "Simulation of genetic systems by automatic digital computers I. Introduction". In: *Australian Journal of Biological Sciences* 10.4, pp. 484–491.

Freund, Yoav and Robert E Schapire (1997). "A decision-theoretic generalization of on-line learning and an application to boosting". In: *Journal of computer and system sciences* 55.1, pp. 119–139.

Friedman, Milton (1937). "The use of ranks to avoid the assumption of normality implicit in the analysis of variance". In: *Journal of the american statistical association* 32.200, pp. 675–701.

Fu, Xin, Chengkai Zhang, Xiang Peng, Lihua Jian, and Zheng Liu (2019). "Towards end-to-end pulsed eddy current classification and regression with CNN". In: *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, pp. 1–5.

Ghorai, Koel, Sourav Saha, Aishwarya Bakshi, Ambuj Mahanti, and Pradeep Ray (2013). "An mhealth recommender for smoking cessation using case based reasoning". In: *2013 46th Hawaii International Conference on System Sciences*. IEEE, pp. 2695–2704.

Gorber, Sarah Connor, Sean Schofield-Hurwitz, Jill Hardt, Geneviève Levasseur, and Mark Tremblay (2009). "The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status". In: *Nicotine & tobacco research* 11.1, pp. 12–24.

Grainge, Matthew J, Lion Shahab, David Hammond, Richard J O'Connor, and Ann McNeill (2009). "First cigarette on waking and time of day as predictors of puffing behaviour in UK adult smokers". In: *Drug and alcohol dependence* 101.3, pp. 191–195.

Gutkin and Serge H Ahmed (2011). *Computational neuroscience of drug addiction*. Vol. 10. Springer Science & Business Media.

Gutkin, Stanislas Dehaene, and Jean-Pierre Changeux (2006). "A neurocomputational hypothesis for nicotine addiction". In: *Proceedings of the National Academy of Sciences* 103.4, pp. 1106–1111.

Haskins, Brianna L, Donna Lesperance, Patric Gibbons, and Edwin D Boudreaux (2017). "A systematic review of smartphone applications for smoking cessation". In: *Translational behavioral medicine* 7.2, pp. 292–299.

Hébert, Emily T, Elise M Stevens, Summer G Frank, Darla E Kendzor, David W Wetter, Michael J Zvolensky, Julia D Buckner, and Michael S Businelle (2018). "An ecological momentary intervention for smoking cessation: the associations of just-in-time, tailored messages with lapse risk factors". In: *Addictive behaviors* 78, pp. 30–35.

Hinton, Geoffrey E, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov (2012). "Improving neural networks by preventing co-adaptation of feature detectors". In: *arXiv preprint arXiv:1207.0580*.

Hoeppner, Bettina B, Susanne S Hoeppner, Lourah Seaboyer, Melissa R Schick, Gwyneth WY Wu, Brandon G Bergman, and John F Kelly (2016). "How smart are smartphone apps for smoking cessation? A content analysis". In: *Nicotine & Tobacco Research* 18.5, pp. 1025–1031.

Holland, JH (1975). "Adaptation in Natural and Artificial Systems, Mich. Univ. Press". In: *Ann Arbor, MI, USA*.

Holst, Arne (2019). *Smartphone users worldwide 2016-2021*.

Huang, Chiou-Jye and Ping-Huan Kuo (2018). "A deep cnn-lstm model for particulate matter (PM2. 5) forecasting in smart cities". In: *Sensors* 18.7, p. 2220.

Huang, Chiou-Jye, Yamin Shen, Yung-Hsiang Chen, and Hsin-Chuan Chen (2020). "A novel hybrid deep neural network model for short-term electricity price forecasting". In: *International Journal of Energy Research*.

Huang, Zili, Shuai Wang, and Kai Yu (2018). "Angular Softmax for Short-Duration Text-independent Speaker Verification." In: *Interspeech*, pp. 3623–3627.

Hughes, John M (2010). *Real World Instrumentation with Python: Automated Data Acquisition and Control Systems*. O'Reilly Media, Inc.

Hughes, John R, Saul Shiffman, Shelly Naud, and Erica N Peters (2017). "Day-to-day variability in self-reported cigarettes per day". In: *Nicotine & Tobacco Research* 19.9, pp. 1107–1111.

Hutter, Frank, Lars Kotthoff, and Joaquin Vanschoren (2019). *Automated machine learning: methods, systems, challenges*. Springer Nature.

Ikram, Sumaiya Thaseen and Aswani Kumar Cherukuri (2016). "Improving accuracy of intrusion detection model using PCA and optimized SVM". In: *Journal of computing and information technology* 24.2, pp. 133–148.

Ioffe, Sergey and Christian Szegedy (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *arXiv preprint arXiv:1502.03167*.

Jeni, László A, Jeffrey F Cohn, and Fernando De La Torre (2013). "Facing imbalanced data–recommendations for the use of performance metrics". In: *2013 Humaine association conference on affective computing and intelligent interaction*. IEEE, pp. 245–251.

Kalantar, Bahareh, Biswajeet Pradhan, Seyed Amir Naghibi, Alireza Motevalli, and Shattri Mansor (2018). "Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)". In: *Geomatics, Natural Hazards and Risk* 9.1, pp. 49–69.

Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom (2014). "A convolutional neural network for modelling sentences". In: *arXiv preprint arXiv:1404.2188*.

Kim, Seong-Hoon, Zong Woo Geem, and Gi-Tae Han (2020). "Hyperparameter Optimization Method Based on Harmony Search Algorithm to Improve Performance of 1D CNN Human Respiration Pattern Recognition System". In: *Sensors* 20.13, p. 3697.

Kim, Tae-Young and Sung-Bae Cho (2018). "Web traffic anomaly detection using C-LSTM neural networks". In: *Expert Systems with Applications* 106, pp. 66–76.

— (2019). "Predicting residential energy consumption using CNN-LSTM neural networks". In: *Energy* 182, pp. 72–81.

Kirchner, Thomas R, Jennifer Cantrell, Andrew Anesetti-Rothermel, Ollie Ganz, Donna M Vallone, and David B Abrams (2013). "Geospatial exposure to point-of-sale tobacco: real-time craving and smoking-cessation outcomes". In: *American journal of preventive medicine* 45.4, pp. 379–385.

Koslovsky, Matthew D, Emily T Hébert, Michael D Swartz, Wenyaw Chan, Luis Leon-Novelo, Anna V Wilkinson, Darla E Kendzor, and Michael S Businelle (2018a). "The time-varying relations between risk factors and smoking before and after a quit attempt". In: *Nicotine and Tobacco Research* 20.10, pp. 1231–1236.

Koslovsky, Matthew D, Michael D Swartz, Wenyaw Chan, Luis Leon-Novelo, Anna V Wilkinson, Darla E Kendzor, and Michael S Businelle (2018b). "Bayesian variable selection for multistate Markov models with interval-censored data in an ecological momentary assessment study of smoking cessation". In: *Biometrics* 74.2, pp. 636–644.

LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

Liu, Tianyuan, Jinsong Bao, Junliang Wang, and Yiming Zhang (2018). "A hybrid CNN–LSTM algorithm for online defect recognition of CO2 welding". In: *Sensors* 18.12, p. 4369.

Long, Xiaojian and Yuchun Wu (2012). "Application of decision tree in student achievement evaluation". In: *2012 International Conference on Computer Science and Electronics Engineering*. Vol. 2. IEEE, pp. 243–247.

Lu, Wenxing, Haidong Rui, Changyong Liang, Li Jiang, Shuping Zhao, and Keqing Li (2020). "A Method Based on GA-CNN-LSTM for Daily Tourist Flow Prediction at Scenic Spots". In: *Entropy* 22.3, p. 261.

Lynch, Krystal L, Jenny E Twesten, Alexandra Stern, and Erik M Augustson (2019). "Level of Alcohol Consumption and Successful Smoking Cessation". In: *Nicotine and Tobacco Research* 21.8, pp. 1058–1064.

Mak, Kwok Kei, Kounseok Lee, and Cheolyong Park (2019). "Applications of machine learning in addiction studies: A systematic review". In: *Psychiatry research*.

Mathieu-Kia, A, S Kellogg, E Butelman, and M Kreek (2002). "Nicotine addiction: insights from recent animal studies". In: *Psychopharmacology* 162.2, pp. 102–118.

McClure, Jennifer B, Andrea L Hartzler, and Sheryl L Catz (2016). "Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers". In: *JMIR mHealth and uHealth* 4.1, e17.

Meenal, R and A Immanuel Selvakumar (2018). "Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters". In: *Renewable Energy* 121, pp. 324–343.

Meng, Erhao, Shengzhi Huang, Qiang Huang, Wei Fang, Lianzhou Wu, and Lu Wang (2019). "A robust method for non-stationary streamflow prediction based on improved EMD-SVM model". In: *Journal of hydrology* 568, pp. 462–478.

Metin, Selin and N Serap Sengor (2012). "From occasional choices to inevitable musts: A computational model of nicotine addiction". In: *Computational intelligence and neuroscience* 2012.

Michie, Susan, Michelle Richardson, Marie Johnston, Charles Abraham, Jill Francis, Wendy Hardeman, Martin P Eccles, James Cane, and Caroline E Wood (2013). "The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions". In: *Annals of behavioral medicine* 46.1, pp. 81–95.

Mim, Mahbina Akter and KS Zamil (2018). "Gis-based analysis of changing surface water in rajshahi city corporation area using support vector machine (svm) decision tree & random forest technique". In: *Machine Learning Research* 3.2, p. 11.

Mobasheri, Mohammad H, Maximilian Johnston, Dominic King, Daniel Leff, Paul Thiruchelvam, and Ara Darzi (2014). "Smartphone breast applications–What's the evidence?" In: *The Breast* 23.5, pp. 683–689.

Moon, Sung Seek, Suk-Young Kang, Weerawat Jitpitaklert, and Seoung Bum Kim (2012). "Decision tree models for characterizing smoking patterns of older adults". In: *Expert Systems with Applications* 39.1, pp. 445–451.

Naughton, Felix, Sarah Hopewell, Neal Lathia, Rik Schalbroeck, Chloë Brown, Cecilia Mascolo, Andy McEwen, and Stephen Sutton (2016). "A context-sensing mobile phone app (Q sense) for smoking cessation: a mixed-methods study". In: *JMIR mHealth and uHealth* 4.3, e106.

Nguyen, Thien Huu and Ralph Grishman (2015). "Relation extraction: Perspective from convolutional neural networks". In: *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pp. 39–48.

NHS (2018). *Coping with cravings -Quit smoking.* `https://www.nhs.uk/live-well/quit-smoking/stop-smoking-coping-with-cravings/`.

Niu, Xiao-Xiao and Ching Y Suen (2012). "A novel hybrid CNN–SVM classifier for recognizing handwritten digits". In: *Pattern Recognition* 45.4, pp. 1318–1325.

ONS (2018). "Adult smoking habits in the UK: 2017". In:

Ordóñez, Francisco Javier and Daniel Roggen (2016). "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition". In: *Sensors* 16.1, p. 115.

Paige, Samantha R, Julia M Alber, Michael L Stellefson, and Janice L Krieger (2018). "Missing the mark for patient engagement: mHealth literacy strategies and behavior change processes in smoking cessation apps". In: *Patient education and counseling* 101.5, pp. 951–955.

Perkins, Kenneth A, Nancy C Jao, and Joshua L Karelitz (2013). "Consistency of daily cigarette smoking amount in dependent adults." In: *Psychology of Addictive Behaviors* 27.3, p. 723.

Phan, Huy, Lars Hertel, Marco Maass, and Alfred Mertins (2016). "Robust audio event recognition with 1-max pooling convolutional neural networks". In: *arXiv preprint arXiv:1604.06338*.

Popa, Dan, Florin Pop, Cristina Serbanescu, and Aniello Castiglione (2019). "Deep learning model for home automation and energy reduction in a smart home environment platform". In: *Neural Computing and Applications* 31.5, pp. 1317–1337.

Pulantara, I Wayan, Bambang Parmanto, and Anne Germain (2018). "Development of a Just-in-Time adaptive mHealth intervention for insomnia: usability study". In: *JMIR human factors* 5.2, e21.

Quinlan, JR (1986). "Induction of Decision Trees. Mach. Learn". In:

— (1993). "The Morgan Kaufmann series in machine learning". In: *San Mateo*.

Rao, Guozheng, Weihang Huang, Zhiyong Feng, and Qiong Cong (2018). "LSTM with sentence representations for document-level sentiment classification". In: *Neurocomputing* 308, pp. 49–57.

Ray, Susmita (2019). "A quick review of machine learning algorithms". In: *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*. IEEE, pp. 35–39.

Regmi, Kabindra, Norhayati Kassim, Norhayati Ahmad, and Nik A Tuah (2017). "Effectiveness of mobile apps for smoking cessation: a review". In: *Tobacco Prevention & Cessation* 3.

Rodgers, Anthony, Tim Corbett, Dale Bramley, Tania Riddell, Mary Wills, Ruey-Bin Lin, and Mark Jones (2005). "Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging". In: *Tobacco control* 14.4, pp. 255–261.

Rosser, Benjamin A and Christopher Eccleston (2011). "Smartphone applications for pain management". In: *Journal of telemedicine and telecare* 17.6, pp. 308–312.

Rumelhart, David E, Geoffrey E Hinton, and Ronald J Williams (1986). "Learning representations by back-propagating errors". In: *nature* 323.6088, pp. 533–536.

Sagheer, Alaa and Mostafa Kotb (2019). "Time series forecasting of petroleum production using deep LSTM recurrent networks". In: *Neurocomputing* 323, pp. 203–213.

Satu, Md Shahriare, Syeda Tanjila Atik, and Mohammad Ali Moni (2020). "A novel hybrid machine learning model to predict diabetes mellitus". In: *Proceedings of International Joint Conference on Computational Intelligence*. Springer, pp. 453–465.

Schick, Robert S, Thomas W Kelsey, John Marston, Kay Samson, and Gerald W Humphris (2018). "MapMySmoke: feasibility of a new quit cigarette smoking

mobile phone application using integrated geo-positioning technology, and motivational messaging within a primary care setting". In: *Pilot and feasibility studies* 4.1, p. 19.

Schuster, Mike and Kuldip K Paliwal (1997). "Bidirectional recurrent neural networks". In: *IEEE transactions on Signal Processing* 45.11, pp. 2673–2681.

Sharma, Kirti and Rainu Nandal (2019). "A Literature Study On Machine Learning Fusion With IOT". In: *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, pp. 1440–1445.

Shiffman, Saul, Michael S Dunbar, Xiaoxue Li, Sarah M Scholl, Hilary A Tindle, Stewart J Anderson, and Stuart G Ferguson (2014). "Smoking patterns and stimulus control in intermittent and daily smokers". In: *PloS one* 9.3.

Shiffman, Saul, Chad J Gwaltney, Mark H Balabanis, Kenneth S Liu, Jean A Paty, Jon D Kassel, Mary Hickcox, and Maryann Gnys (2009). "Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment." In:

Shiffman, Saul, Michael Hufford, Mary Hickcox, Jean A Paty, Maryann Gnys, and Jon D Kassel (1997). "Remember that? A comparison of real-time versus retrospective recall of smoking lapses." In: *Journal of consulting and clinical psychology* 65.2, p. 292.

Solomon, Richard L (1980). "Recent experiments testing an opponent-process theory of acquired motivation". In: *Acta Neurobiol Exp (Wars)* 40.1, pp. 271–289.

Sorokin, Maksim, A Zhdanov, and D Zhdanov (2018). "Recovery of optical parameters of a scene using fully-convolutional neural networks". In: *Proc. MICSECS*.

Stevenson, Jennifer G, Jason A Oliver, Matthew B Hallyburton, Maggie M Sweitzer, Cynthia A Conklin, and F Joseph McClernon (2017). "Smoking environment cues reduce ability to resist smoking as measured by a delay to smoking task". In: *Addictive behaviors* 67, pp. 49–52.

Sukawattanavijit, Chanika, Jie Chen, and Hongsheng Zhang (2017). "GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data". In: *IEEE Geoscience and Remote Sensing Letters* 14.3, pp. 284–288.

Sun, Yanan, Bing Xue, Mengjie Zhang, Gary G Yen, and Jiancheng Lv (2020). "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification". In: *IEEE Transactions on Cybernetics*.

Tang, Yichuan (2013). "Deep learning using linear support vector machines". In: *arXiv preprint arXiv:1306.0239*.

Tao, Peiying, Zhe Sun, and Zhixin Sun (2018). "An improved intrusion detection algorithm based on GA and SVM". In: *Ieee Access* 6, pp. 13624–13631.

Tao, Zhou, Lu Huiling, Wang Wenwen, and Yong Xia (2019). "GA-SVM based feature selection and parameter optimization in hospitalization expense modeling". In: *Applied Soft Computing* 75, pp. 323–332.

Tavakoli, Neda (2019). "Modeling genome data using bidirectional LSTM". In: *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*. Vol. 2. IEEE, pp. 183–188.

Thaseen, Ikram Sumaiya and Cherukuri Aswani Kumar (2017). "Intrusion detection model using fusion of chi-square feature selection and multi class SVM". In: *Journal of King Saud University-Computer and Information Sciences* 29.4, pp. 462–472.

Tian, Yan, Kaili Zhang, Jianyuan Li, Xianxuan Lin, and Bailin Yang (2018). "LSTM-based traffic flow prediction with missing data". In: *Neurocomputing* 318, pp. 297–305.

Tregarthen, Jenna P, James Lock, and Alison M Darcy (2015). "Development of a smartphone application for eating disorder self-monitoring". In: *International Journal of Eating Disorders* 48.7, pp. 972–982.

Treloar, Hayley R, Thomas M Piasecki, Danielle E McCarthy, and Timothy B Baker (2014). "Relations among caffeine consumption, smoking, smoking urge, and subjective smoking reinforcement in daily life". In: *Journal of Caffeine Research* 4.3, pp. 93–99.

Tu, My Chau, Dongil Shin, and Dongkyoo Shin (2009). "A comparative study of medical data classification methods based on decision tree and bagging algorithms". In: *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*. IEEE, pp. 183–187.

Ubhi, Harveen Kaur, Susan Michie, Daniel Kotz, Onno CP van Schayck, Abiram Selladurai, and Robert West (2016). "Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features". In: *Translational behavioral medicine* 6.3, pp. 410–417.

Ubhi, Harveen Kaur, Susan Michie, Daniel Kotz, Wai Chi Wong, and Robert West (2015). "A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28". In: *Journal of medical Internet research* 17.1, e17.

Vanli, N Denizcan, Muhammed O Sayin, Mohammadreza Mohaghegh, Huseyin Ozkan, and Suleyman S Kozat (2019). "Nonlinear regression via incremental decision trees". In: *Pattern Recognition* 86, pp. 1–13.

Vapnik, Vladimir (2013). *The nature of statistical learning theory*. Springer science & business media.

Verma, Abhishek and Virender Ranga (2020). "Machine learning based intrusion detection systems for IoT applications". In: *Wireless Personal Communications* 111.4, pp. 2287–2310.

Vilardaga, Roger, Elisabet Casellas-Pujol, Joseph F McClernon, and Kathleen A Garrison (2019). "Mobile applications for the treatment of tobacco use and dependence". In: *Current Addiction Reports* 6.2, pp. 86–97.

Walsh, Jane C and Jenny M Groarke (2019). "Integrating behavioral science with mobile (mhealth) technology to optimize health behavior change interventions". In: *European Psychologist*.

Wang, Fei, Zhao Zhen, Bo Wang, and Zengqiang Mi (2018). "Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting". In: *Applied Sciences* 8.1, p. 28.

Wang, Yi, Dahua Gan, Mingyang Sun, Ning Zhang, Zongxiang Lu, and Chongqing Kang (2019). "Probabilistic individual load forecasting using pinball loss guided LSTM". In: *Applied Energy* 235, pp. 10–20.

Wardhani, Ni Wayan Surya, Masithoh Yessi Rochayani, Atiek Iriany, Agus Dwi Sulistyono, and Prayudi Lestantyo (2019). "Cross-validation Metrics for Evaluating Classification Performance on Imbalanced Data". In: *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*. IEEE, pp. 14–18.

Webb, Jamie, Sarrah Peerbux, Peter Smittenaar, Sarim Siddiqui, Yusuf Sherwani, Maroof Ahmed, Hannah MacRae, Hannah Puri, Sangita Bhalla, and Azeem Majeed (2020). "Preliminary Outcomes of a Digital Therapeutic Intervention for Smoking Cessation in Adult Smokers: Randomized Controlled Trial". In: *JMIR mental health* 7.10, e22833.

Webb, Thomas L, Betty PI Chang, and Yael Benn (2013). "'The Ostrich Problem': Motivated avoidance or rejection of information about goal progress". In: *Social and Personality Psychology Compass* 7.11, pp. 794–807.

Williams, Drew, Golam Mushih Tanimul Ahsan, Ivor Addo, Sheikh Iqbal Ahamed, Daniel Petereit, Linda Burhansstipanov, Linda Krebs, and Mark Dignan (2018). "Building a Tailored Text Messaging System for Smoking Cessation in Native

American Populations". In: *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*. Vol. 1. IEEE, pp. 866–874.

Wu, Qingyao, Mingkui Tan, Hengjie Song, Jian Chen, and Michael K Ng (2016). "ML-FOREST: A multi-label tree ensemble method for multi-label classification". In: *IEEE transactions on knowledge and data engineering* 28.10, pp. 2665–2680.

Wu, Yu-Xi, Qing-Biao Wu, and Jia-Qi Zhu (2019). "Data-driven wind speed forecasting using deep feature extraction and LSTM". In: *IET Renewable Power Generation* 13.12, pp. 2062–2069.

Xiong, Zheng, Yuxin Cui, Zhonghao Liu, Yong Zhao, Ming Hu, and Jianjun Hu (2020). "Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation". In: *Computational Materials Science* 171, p. 109203.

Xue, Di-Xiu, Rong Zhang, Hui Feng, and Ya-Lei Wang (2016). "CNN-SVM for microvascular morphological type recognition with data augmentation". In: *Journal of medical and biological engineering* 36.6, pp. 755–764.

Yan, Yan, Ying Wang, Wen-Chao Gao, Bo-Wen Zhang, Chun Yang, and Xu-Cheng Yin (2018). "LSTM2: Multi-Label Ranking for Document Classification". In: *Neural Processing Letters* 47.1, pp. 117–138.

Zahid, Maheen, Fahad Ahmed, Nadeem Javaid, Raza Abid Abbasi, Hafiza Syeda Zainab Kazmi, Atia Javaid, Muhammad Bilal, Mariam Akbar, and Manzoor Ilahi (2019). "Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids". In: *Electronics* 8.2, p. 122.

Zhang, Xiao, Yaodong Yu, Lingxiao Wang, and Quanquan Gu (2019a). "Learning one-hidden-layer relu networks via gradient descent". In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 1524–1534.

Zhang, Yupu, Jinhai Liu, Zhihang Zhang, and Junnan Huang (2019b). "Prediction of Daily Smoking Behavior Based on Decision Tree Machine Learning Algorithm". In: *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*. IEEE, pp. 330–333.

Zhao, Rui, Ruqiang Yan, Jinjiang Wang, and Kezhi Mao (2017). "Learning to monitor machine health with convolutional bi-directional LSTM networks". In: *Sensors* 17.2, p. 273.

Zhou, Chunting, Chonglin Sun, Zhiyuan Liu, and Francis Lau (2015). "A C-LSTM neural network for text classification". In: *arXiv preprint arXiv:1511.08630*.

Zhou, Xinjie, Xiaojun Wan, and Jianguo Xiao (2016). "Attention-based LSTM network for cross-lingual sentiment classification". In: *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 247–256.

Zhu, Wei, Mei Xie, and Jian-Feng Xie (2012). "A decision tree algorithm for license plate recognition based on bagging". In: *2012 International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP)*. IEEE, pp. 136–139.

Zidi, Salah, Tarek Moulahi, and Bechir Alaya (2017). "Fault detection in wireless sensor networks through SVM classifier". In: *IEEE Sensors Journal* 18.1, pp. 340–347.

Zou, Difan, Yuan Cao, Dongruo Zhou, and Quanquan Gu (2020). "Gradient descent optimizes over-parameterized deep ReLU networks". In: *Machine Learning* 109.3, pp. 467–492.

Appendix A

Published papers

2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation

Decision Tree Model of Smoking Behaviour

Maryam Abo-Tabik
Department of Computing and Mathematics
Faculty of Science and Engineering
Manchester Metropolitan University
Manchester, United Kingdom
maryam.a.abo-tabik@stu.mmu.ac.uk

Nicholas Costen
Department of Computing and Mathematics
Faculty of Science and Engineering
Manchester Metropolitan University
Manchester, United Kingdom
N.Costen@mmu.ac.uk

John Darby
Department of Computing and Mathematics
Faculty of Science and Engineering
Manchester Metropolitan University
Manchester, United Kingdom
J.Darby@mmu.ac.uk

Yael Benn
Department of Psychology
Manchester Metropolitan University
Manchester, United Kingdom
Y.Benn@mmu.ac.uk

***Abstract*—Smoking is considered the cause of many health problems. While most smokers wish to quit smoking, many relapse. In order to support an efficient and timely delivery of intervention for those wishing to quit smoking, it is important to be able to model the smoker's behaviour. This research describes the creation of a combined Control Theory and Decision Tree Model that can learn the smoker's daily routine and predict smoking events. The model structure combines a Control Theory model of smoking with a Bagged Decision Tree classifier to adapt to individual differences between smokers, and predict smoking actions based on internal stressors (nicotine level, withdrawal, and time since the last dose) and external stressors (e.g. location, environment, etc.). The designed model has 91.075% overall accuracy of classification rate and the error rate of forecasting the nicotine effect using the designed model is also low (MSE=0.048771, RMSE=0.216324, and NRMSE=0.153946) for regular days and (MSE=0.048804, RMSE=0.216637, and NRMSE=0.195929).**

***Index Terms*—smoker's behaviour, addictive behaviour, machine learning, Decision Tree, Bagged Decision Tree, Control Theory.**

I. Introduction

Smoking is considered one of the leading causes of deaths internationally. According to a recent NHS report [1] in 2016, smoking caused the death of about 77,900 people in England alone. The report further states that smoking is not only harmful to the smokers, but many diseases might be caused by the exposure to passive smoking, especially affecting children who are particularly vulnerable to the effects of passive smoking. This makes reducing cigarette smoking a public health priority.

Actions (including smoking) can be seen as being motivated by the need of the human system to maintain stability, over a range of time-scales, in the face of a changing environment. This motivation can appear in the form of internal feelings such as sadness, or external need such as maintaining nicotine level [2]. Closed-loop control model is a common instrumental modelling method that seeks to maintain stability. It employs the feedback principle, useing the output data from the model (feedback signal) as an input to modify the model's actions, and hence maintain stability [3]. However, modelling addictive behaviour as a closed loop control model is a challenging task. It requires understanding the complexity of humans, as well as determining what elements should be counted to model the addictive behaviour. Moreover, when modelling the addictive behaviour, the goal state represents the fact that the system seeks to obtain a steady state (natural state), rather than to imply that there exists a single fixed value, as is often the case in system engineering [4].

Opponent process theory is claimed to be an essential method that can be used to model a person's emotional state [5]. Solomon [6] described addictive behaviour using the opponent process theory. Within this model, an addict experiences pleasure as soon as a drug is supplied, which is followed by slowly accumulated withdrawal symptoms. As such, during the initial stages of addiction, the pleasure level is high and is accompanied by a low level of withdrawal symptoms. However, as time goes by, the withdrawal symptoms increase leading to a decrease in pleasure caused by using the drug, potentially resulting in a higher quantity of the drug being consumed [4].

Bobashev et al. [7] modelled the behaviour of smokers and employed the opponent process scheme of control theory. The model did not present any complex neurobiological process, only providing a mathematical model with a cascading feedback loop, aimed at presenting the scientific narrative of the opponent process as shown in Fig. 1.

The model equations were developed with phenomenological interpretation in mind, and no real biological process was modelled. A set of continuous functions were used feed into the cascading functions. The system equations involve five interlinked processes,

$$ProcessA : \frac{dY_1}{dt} = e^{-\alpha t} - b_1 Y_1 \quad (1)$$

978-1-7281-4034-6/19/$31.00 ©2019 IEEE
DOI 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00311

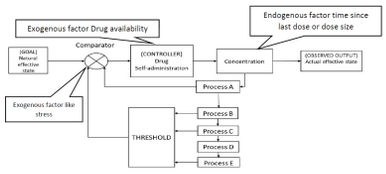


Fig. 1: Control theory model of smoking based on [7].

$$ProcessB: \frac{dY_2}{dt} = a_1Y_1 - b_2Y_2 \tag{2}$$

$$ProcessC: \frac{dY_3}{dt} = a_2Y_2 - b_3Y_3 \tag{3}$$

$$ProcessD: \frac{dY_4}{dt} = a_3Y_3 - b_4Y_4 \tag{4}$$

$$ProcessE: \frac{dY_5}{dt} = a_4Y_4 - b_5Y_5 \tag{5}$$

where a, b and α are scaling coefficients, and all the Y_i initial values are equal set to zero. Each equation presents a weighted integration of the previous one, causing the processes to lengthen sucessively. Y_1 represented the effect of nicotine level and is modelled with a pharmacokinetic equation. Y_2 represents the toxicity level and how the body processes the drug. Y_3 is the daily smoking habit. Y_5 is a longer scaling habit, which is scaled in years (rather than minutes/ hours/ days). While the process Y_4 has not been interpreted, it has been used to add scaling period between Y_3 and Y_5, which results in a slow change in process Y_5. To simulate smoking behaviour, a threshold value was defined to prompt self-administration. The threshold

$$T = \frac{(\beta_3Y_3 + \beta_5Y_5)}{(1 + \beta_2Y2)} \tag{6}$$

has calibration coefficients β_i, and to avoid division by zero one is added to the denominator of the equation. The threshold value i s changed based on external stressors to initiate cigarette use

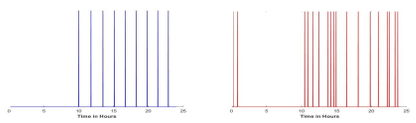
$$T = T + stress. \tag{7}$$

The research also modelled the withdrawal and craving processes; these p rocesses begin immediately following the initial nicotine use and grow over time

$$W = \frac{d_3Y_3(T - Y_1)}{(Y_{0w} + Y_1)} \tag{8}$$

$$C = \frac{d_5Y_5(T - Y_1)}{(Y_{0c} + Y_1)} \tag{9}$$

where d_3, d_5, Y_{0w} and Y_{0c} are calibration coefficients. This control theory model was able to simulate the changes in smoking behaviour over time. However, the system was not able to present real-life behaviour, and could not capture individual differences between smokers' daily habits. Fig. 2 presents the differences between the smoking behaviour as presented using the simulated control theory model Fig. 2a and real-life data collected from a participant shown in Fig. 2b.



(a) Simulated smoking events (b) Real smoking events

Fig. 2: Smoking frequency; each peak represents a smoking event (a) a simulated smoking behaviour generated by the control theory model [7] , and (b) real smoking behaviour on a randomly-selected day from our collected data.

Studies show that modelling smoking behaviour is essential, it can improve the intervention process in the way of helping smokers in their most needed time [8]. While most of the known approaches try to find a relationship between some clues (e.g., withdrawal, stress, place, and the presence of other smokers) and urge to smoke. Most of these studies rely on participants self-reporting these indicators, as the results indicated that these predictors provide a high degree of possibility for predicting potential smoking events or relapse in quitting period. However, Self-reporting as a method can be inaccurate as it is sensitive to self-biased errors [9]. Another research [10] investigates the possibility of using hidden Markov models to set patterns for the timing and places that the smokers are most likely to smoke, and then use these patterns for better delivery of the support messages. The paper did not report any analytical result that is related to Hidden Markov models, except the positive feedback from the participants who used their mobile application.

As such, the current research aims to develop a machine learning model, which when combined with a control theory model of smoking, will be able to adapt to the smoker's unique behaviour and predict future smoking events. The Bobashev et al. [7] model was chosen due to its ability to capture the nicotine effect using the pharmacokinetic equation. Here, we describe the implementation of this control theory model of smoking that is expanded to incorporate other factors affecting smokers' smoking behaviour (e.g., location and activity).

II. Decision Tree for classifying unimodal tabular data

Many classification problems have a large dataset containing complex information, including potential labelling inaccuracies. A decision tree is considered to be an efficient machine learning classifiers for such problems [11]. An early version of the regression tree is the classification and regression tree CART[12]; it recursively divided the dataset based on the

selected features using Least Squared Deviation (LSD) as its impurity function [13],

$$R(t) = \frac{1}{N_w(t)} \sum_{i \in t} w_i f_i (y_i - \bar{y}_i(t))^2 \quad (10)$$

$$\bar{y}(t) = \frac{1}{N_w(t)} \sum_{i \in t} w_i f_i y_i \quad (11)$$

$$N_w(t) = \sum_{i \in t} w_i f_i \quad (12)$$

where $N_w(t)$ is the weighted number of samples in node t, w_i is the calculated weight value for each i and f_i is the recorded response. y_i is the response and $\bar{y}_i$ is the value of the mean response. t The splitting process is performed using

$$Q(s,t) = R(t) - R(t_L) - R(t_R) \quad (13)$$

where t_L is the left child and t_R is the right child of the node t.

Many classifications enhanced their models by training their dataset using several classifiers, and the results are then combined using a voting process, this general method being called an ensemble classifier [14]. The ensemble has also been used with decision trees, mainly in two approaches; either Bagging [15], or boosting [16] algorithms. Bagging (or bootstrap aggregating) is applied to decision trees by generating multiple versions of decision trees during the training process and using a plurality vote between them to predict the class. The idea is to create several subsets from the dataset, with each subset training its own decision tree, and then combine the result from several trained models in order to reach a more reliable predictor and reduce the variance of classification [17, 18]. Boosting is the use of iterative re-training, so as to create the ensemble sequentially, where at each step the later trained classifier is learning from the previous errors generated by the earlier classifiers, by increasing the weight as the training progresses [19]. While boosting classifiers increases the accuracy of the trained model over bagging, in return it increases the chance of overfitting; another drawback for boosting is that it is very slow, and it is sensitive to noise [20].

Another form of the ensembles decision tree is the random forest; this model is efficient because it reduces the overfitting problem [21]. Random forest randomly selects subset samples from the training set (in-bagging) and use them to generate multiple versions of the decision tree. The rest of the samples (out-bagging) will be used in cross-validation to estimate how well the classifier works. The generated error from the validation process is called out-of-bag (OOB) error. Random forest is automatically produced without any pruning, and each node splits using a predefined number of features. The forest grows up to a set limit of the number of trees. Random forest generates trees with low bias and high variance. The classification output is calculated by averaging the class assignment probability generated by all the trees; the probability of the class is calculated using all the produced trees [22].

III. Data collection and processing

There is currently no published dataset that can fit the needs of our research. Moreover, to create a data set that can be employed in modelling smoking behaviour, several steps were followed. A mobile application was used to collect signals from mobile sensors (e.g., movement and environment) for approximately two weeks, while users reported their smoking events. Three types of events occurr in the dataset, which are labelled as smoking (1), not smoking (2) and app-off (0) events. The later was labelled as app-off due to gaps in the dataset (i.e. the participant's mobile phone was off). Table I shows the frequency of events for each of the four participants. One problem that can be seen is that the classes are unbalanced, as the number of smoking events is much lower than the number of non-smoking events. Overall, there are 1440 data samples per day (one sample per minute), while the reported smoking events are less than 15 per day, and the rest are either not smoking or app-off events. To overcome this problem the model is targeting at the smoking period, not at the per-minute smoking event. Instead, the data labeling changed to include a 10-minute window, hence reducing the ratio of smoking to non-smoking events. Table II shows the frequency of events for each of the four participants after applying the change.

TABLE I: The number of labels in each of the three labelling categories.

	App off	**Smoking**	**Not smoking**
Participant 1	451	201	18068
Participant 2	6307	64	12349
Participant 3	3997	66	14657
Participant 4	15514	82	3124

TABLE II: The number of labels in each of the three labelling categories after applying a 10 minute smoking window.

	App off	**Smoking**	**Not smoking**
Participant 1	451	1960	16308
Participant 2	6217	630	11872
Participant 3	3997	650	14072
Participant 4	15211	800	2708

The reported smoking events are then used as input to the control theory model of smoking, in order to calculate the nicotine levels and threshold value during the 13 days. One 24 hour period was dropped because it was made of two half-days (one at the start and the other at the end of the data collection period). All the calculated data along with collected mobile data (eg. light, GPS Location, and activity labels etc.) are all combined to form the dataset tables for each participant. The labelled smoking events will be the labels for the data set. Fig. 3 shows the process of data collection.

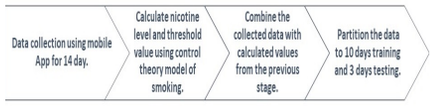
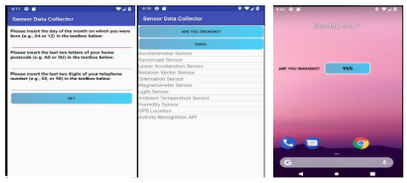


Fig. 3: Overview of the study: data collection and processing steps.

A. Mobile App

Data collection took place using a mobile application developed for Android mobile users, using Android Studio (IDE). The main focus of the User Interface (UI) was to develop a user-friendly interface that provides no feedback to users, as so to avoid influencing their behaviour[23]. The UI was used to label smoking events, relying on participants' self-reporting of events. Users could report smoking events either by pressing a button on the main layout of the App, or by pressing a Widget on the home screen of the smartphone as can be seen in Fig. 4.



(a) Participant registration screen (b) Application main layout (c) Amart phone home screeen

Fig. 4: Mobile application UI

The application was designed to run as a background service, which records data from the phone's sensors. This service was designed to restart itself whenever terminated (either by the OS or otherwise). This was implemented in order to overcome a new restriction forced by Android on the development of background services that run for long periods. The background service recorded one sample per minute from the sensed data. Collected data, along with smoking events were stored on an internal SQLite database.

B. Data collection

For this study, the participants were smoking adult over 18 years old, with a good level of English literacy. They each owned an Android mobile phone. Smokers are defined as those smoking at least 5 cigarettes a day. During the data collection period, the application was installed on the participant's smartphone for two weeks. No restrictions have been placed on their daily activities, and they have only been asked to report their smoking events and keep the GPS on. At this stage of the research data has been collected from 4 participants (3 females:1 male)[1].

Data were collected from several sensors in order to identify correlations between smoking events and the sensors reading. Table III shows the types of collected data. The goal is to use the collected data to find the association between smoking events and environmental data, in order to inform the implementation of a machine learning model that can automatically predict smoking events based on the occurrence of internal and external predictors. Following data collection, it emerged that not all sensors are available in all mobile models. Therefore the plan was modified to use only the common sensors that appear in most of the mobiles, i.e., the accelerometer and light sensors along with GPS values and human activity labels.

TABLE III: The number of labels in each of the three labelling categories.

Collected data group name	Description
ID	This is unique ID that Identify the user data, it is set by the user at the start of the study.
Timing value	This is time stamp DD-MM-YYYY,HH:MM:SS
Motion sensors data	Accelerometer, Gyroscope, Linear acceleration, Orientation, Rotation vector.
Environmental data	Magnetic field, Light level, Ambient temperature, Relative humidity, GPS location.
Activity labels	Google activity recognition API (Still, Running, Walking, Cycling, Tilting, and Driving).
Smoking labels	This is labelled by the user.

IV. Approach to Model Development

To design a machine learning model for smoking behaviour the control theory model of smoking will be combined with the decision tree classifier. At the start, each part of the model will be analysed separately before reaching the final model.

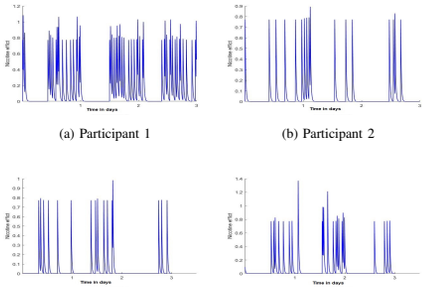
A. Control theory model of smoking

Using the reported smoking events, nicotine concentration was calculated using the control theory model of smoking [7] as shown in Fig. 5. Each peak in the figure represents smoking events, followed by a gradual decrease in the nicotine level until the next smoking event.

Fig. 6 shows the threshold values calculated using the control theory model. The peaks represent no smoking periods, the value of the threshold decreases by the increased number of cigarettes per day.

The control theory model also models the withdrawal and craving symptoms, Fig. 7 shows the values of withdrawal and craving over 10 days period.

[1]Although the number of participants appears small, a large volume of data was collected from each participant (approximately 1010 smoking events and 18720 samples each), making it sufficient for modeling a machine learning problem.



(a) Participant 1 (b) Participant 2

(c) Participant 3 (d) Participant 4

Fig. 5: Examples of 3 days of smoking behaviour by four participants, as modelled using control theory to represent nicotine levels.

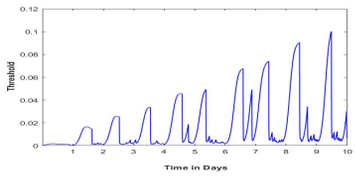
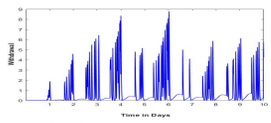
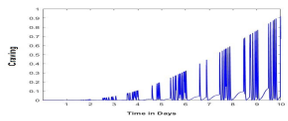


Fig. 6: Example of 10 days calculated threshold value using the control theory model of smoking and collected data from one of the participants.



(a) Withdrawal value based on the control theory model of smoking.



(b) Craving value based on the control theory model of smoking

Fig. 7: Example of 10 days calculated withdrawal and craving values using the control theory model of smoking and collected data from one of the participants.

B. Classification of smoker behavioural data

Three types of events occurr in the collected dataset, which are labelled as smoking (1), not smoking (2) and App app-off (0) events, where the later occur due to gaps in the dataset (e.g. participant turns the mobile off). Three types of Decision Tree models were explored; CART, Boosted Tree, and Tree Bagging (this selects a random subset of samples as in the random forest algorithm). The three classifiers are implemented and tested using the Matlab2017 "Statistics and Machine Learning Toolbox".

Initially, the classification methods are tested to see whether the classifier can detect the smokers events using only endogenous factors. Time, nicotine level, and threshold are used as input to the decision tree classifier. Table IV shows the classification accuracy, The data was tested using the iterative bootstrap process where three users are held for training and validation, and one participant is used for testing. The routine is repeated for each participant.

TABLE IV: The precision level of classification test based on only endogenous factors.

Calculated accuracy category	**The percentage accuracy level**		
	Tree Bagging	***Boosted Tree***	***CADT***
Participant 1 App off	4.7	2.6	1
Participant 1 smoking	70.2	83	65.6
Participant 1 not smoking	94.9	94.2	88.5
Participant 1 overall	54.46	20.8	27.59
Participant 2 App off	31.8	37.3	37
Participant 2 smoking	88	60.9	18.6
Participant 2 not smoking	64	66.9	66
Participant 2 overall	64	55.6	49.73
Participant 3 App off	39.3	17	17.4
Participant 3 smoking	73.1	77	43.7
Participant 3 not smoking	89.2	72.5	72.6
Participant 3 overall	68.28	43.08	42.73
Participant 4 App off	94.3	95.4	91.6
Participant 4 smoking	85.3	76.6	42
Participant 4 not smoking	15.6	16.8	17.7
Participant 4 overall	23.93	29.6	34.59
Average App off	42.525	38.075	36.75
Average smoking	79.15	74.375	42.475
Average not smoking	65.925	62.6	61.2
Average overall	52.6675	37.27	38.66

Secondly, to test the effect of adding the external factors on the performance of the classifier, GPS Location, light level, and human motion label are all used as predictors by the three classification methods along with the endogenous factors. Since the exogenous factors are personalized for each participant, the training model needs to be trained for each participant. The collected dataset for each participant was portioned into 10 days training (70% training and 30% validation) and 3 days testing. Table V shows the result of the testing process.

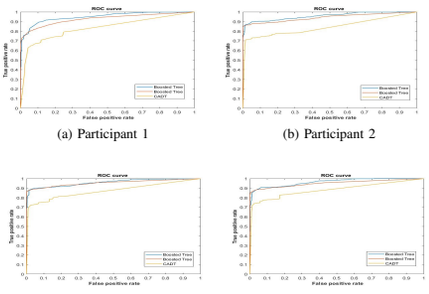
It can see from the tables that in general the performance of the Tree Bagging method is better than the other two classifiers, and that using all 6 predictors can give better overall performance. This can result in the conclusion that in order to model the smoker's behaviour the model has to

TABLE V: The precision level of classification test based all 6 predictors.

Calculated accuracy Category	The percentage accuracy level		
	CADT	*Boosted Tree*	*Tree Bagging*
Participant 1 overall	92.1	68	95.2
Participant 1 smoking	64.4	23.4	87.1
Participant 1 not smoking	97.0	97.7	96.1
Participant 1 unknown	0.0	16.8	87.5
Participant 2 overall	77.1	51.8	73.8
Participant 2 smoking	37.7	3.6	19.1
Participant 2 not smoking	85.5	89.8	95.5
Participant 2 unknown	69.7	64.9	68.1
Participant 3 overall	98.4	75.2	98.8
Participant 3 smoking	68	7.6	97.5
Participant 3 not smoking	99.5	99.8	98.1
Participant 3 unknown	100	97.1	100
Participant 4 overall	82.6	90.6	96.5
Participant 4 smoking	26.2	19.8	79.8
Participant 4 not smoking	19.8	11	61.3
Participant 4 unknown	99.3	96.8	98.8
Average overall	87.55	71.4	91.075
Average smoking	49.075	13.6	70.875
Average not smoking	75.45	74.575	87.75
Average unknown	67.25	68.9	88.6

be trained based on the individual behaviour for each person, and a general model will not target the unique needs that each person may have.

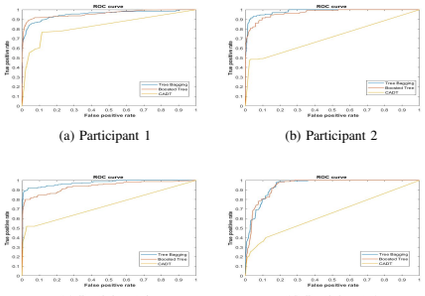
The ROC (Receiver Operating Characteristic) curve clarifies the differences in the performance between the three classifiers and shows how the performance increases when all the predictors are used. Fig. 8 and 9 compare the performance of the classifiers based on the classification methods and the number of input features, where the first figure shows the ROC curve for four participants using only the endogenous factors, while the second figure shows the classification performance for the four participants after considering all 6 predictors.



(a) Participant 1 (b) Participant 2

(c) Participant 3 (d) Participant 4

Fig. 8: Standerd ROC curves for smoking labels classification using only endogenous factors.



(a) Participant 1 (b) Participant 2

(c) Participant 3 (d) Participant 4

Fig. 9: Standerd ROC curves for smoking labels classification using 6 factors.

Other performance measures are displayed in TableVI;it can see from the table that the performance of the Bagging Tree is higher than the other classifiers.

TABLE VI: Performance indices for three classification methods.

Performance index	*Tree Bagging*	*Boosted Tree*	*CADT*
Precision	0.8858	0.7979	0.753735325
Recall	0.8117	0.77087	0.71498
F1 score	0.8282	0.7362	0.7129
Accuracy	0.9142	0.8678	0.8910

The bagging decision tree's ability to minimise the effect of the overfitting problem increased its performance over the other classification methods. Fig.10 shows the out-of-bag error against the number of classification trees grown.

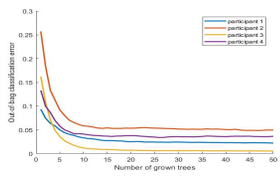


Fig. 10: Out-of-bag error against the number of classification trees grown.

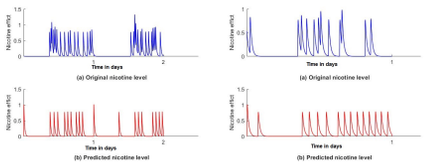
V. RESULTS

After testing the three classification methods, the Bagging Tree method was selected as a classifier to predict smoking events. The classifier predicts either smoking or non-smoking states, with the App off event being treated as non-smoking

events. The point of the prediction to see if it can forecast the nicotine level (other than the original calculated values) using combined control theory and machine learning model.

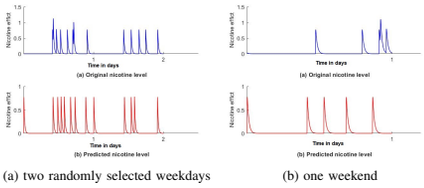
The machine learning model combined with control theory model of smoking to model the smoker's daily behaviour in order to detect the smoking events using endogenous factors, and the other collected data (GPS Location, light level and human motion label). Since the exogenous factors are personalised for each participant, the training model needs to be trained for each participant. The data was tested iteratively, each participant data have been separated for twelve-day training and one-day testing, and then the routine is repeated for each day. This process helped in comparing the prediction level based on different day of the week.

Fig. 11 and 12 shows the prediction result for two participants for randomly selected two regular weekdays along with the prediction of one of the weekend days for the same participant. All 6 predictors were used as input to the system. The nicotine level was predicted during the closed-loop process; no pre-calculated data was used.



(a) two randomly selected weekdays (b) one weekend

Fig. 11: Example of predicted nicotine level for participant 1.



(a) two randomly selected weekdays (b) one weekend

Fig. 12: Example of predicted nicotine level for participant 2.

Although some smoking events were missed, the model in general reliably models the smoking behaviour of each of the participants. The model is strongly relaying on the cooperation from the participants when reporting the smoking events accurately. The final design of model of the daily smoker's behaviour can be seen in Fig 13.

The results of the Mean Square Error (MSE), Root Mean Square Error (RMSE), and Normalized Root Mean Square Error (NRMSE), which are the error criteria used to measure the performance of the model, are displayed Table VII and VIII.

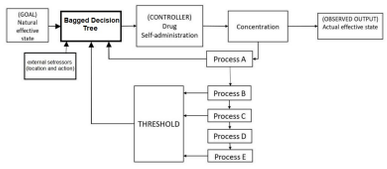


Fig. 13: Smoking behaviour model utilizing machine learning. Data are collected and processed using the steps described in Fig3. The 6 predictors are used as input to the Bagged decision tree classifier. A classification value of 1 represents a potential smoking event. This value is passed to the CONTROLLER, simulating the taking of a cigarette, and re-initializing the parameters of the control model to zero.

TABLE VII: The overall error rate of the proposed model over the regular days.

	MSE	***RMSE***	***NRMSE***
Participant 1	0.082667	0.287519	0.199763
Participant 2	0.038543	0.196323	0.124366
Participant 3	0.045966	0.214396	0.169778
Participant 4	0.027908	0.167057	0.121877
Average	0.048771	0.216324	0.153946

TABLE VIII: The overall error rate of the proposed model over the weekends.

	MSE	***RMSE***	***NRMSE***
Participant 1	0.084701	0.291034	0.202995
Participant 2	0.034426	0.185542	0.167896
Participant 3	0.039972	0.199929	0.203229
Participant 4	0.036116	0.190042	0.209595
Average	0.048804	0.216637	0.195929

VI. Conclusions

In conclusion, machine learning was sucessfully applied to model smokers' behaviour. The design model at this stage combines Bagged Decision Tree with the control theory model of smoking, and the results are generally promising. Six predictors of smokers' behaviour (nicotine effect level, the threshold value as calculated by control theory, light sensor, GPS location and type of activity) have been used to predict the smoking events. This design was able to adapt to the behaviour of individual smokers, but the accuracy of the smoking event prediction can still be improved.

It is expected that the accuracy of the system in predicting the smoking events will be increased by taking advantage of the information such as the indoor smoking ban in the UK

and replacing the Google activity recognition by more accurate human behaviour classifier using the collected accelerometer values. It may also be possible to construct a combined model of individuals' behaviour, using additional external data such as the addresses of their work and home, and also public information on the location of businesses such as bars and resturants likely to be associated with smoking. These additons to the model are currently under consideration.

ACKNOWLEDGMENT

We thank the Manchester Metropolitan Unviersity intern students, employed under the funding program from the faculty of Health, Psychology and Social Care UKRIS and Faculty of Science and Engineering, who collected the data necessary for this study, and also the participants who gave up their time and battery power for this research.

REFERENCES

[1] NHS, "Adult smoking habits in the uk: 2016," *office of national statistics*, June 2018.

[2] M. S. Fibla, U. Bernardet, and P. F. Verschure, "Allostatic control for robot behaviour regulation: An extension to path planning," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 1935–1942.

[3] J. M. Hughes, *Real World Instrumentation with Python: Automated Data Acquisition and Control Systems*. " O'Reilly Media, Inc.", 2010.

[4] B. Gutkin and S. H. Ahmed, *Computational Neuroscience of Drug Addiction*. Springer Science & Business Media, 2011, vol. 10.

[5] S. Metin and N. S. Sengor, "From occasional choices to inevitable musts: A computational model of nicotine addiction," *Computational intelligence and neuroscience*, vol. 2012, p. 18, 2012.

[6] R. L. Solomon, "The opponent-process theory of acquired motivation: the costs of pleasure and the benefits of pain." *American psychologist*, vol. 35, no. 8, p. 691, 1980.

[7] G. Bobashev, J. Holloway, E. Solano, and B. Gutkin, "Control theory model of smoking." *American psychologist*, 2017.

[8] K. P. Timms, D. E. Rivera, L. M. Collins, and M. E. Piper, "Control systems engineering for understanding and optimizing smoking cessation interventions," in *2013 American Control Conference*. IEEE, 2013, pp. 1964–1969.

[9] M. S. Businelle, P. Ma, D. E. Kendzor, S. G. Frank, D. W. Wetter, and D. J. Vidrine, "Using intensive longitudinal data collected via mobile phone to detect imminent lapse in smokers undergoing a scheduled quit attempt," *Journal of medical Internet research*, vol. 18, no. 10, 2016.

[10] R. S. Schick, T. W. Kelsey, J. Marston, K. Samson, and G. W. Humphris, "Mapmysmoke: feasibility of a new quit cigarette smoking mobile phone application using integrated geo-positioning technology, and motivational messaging within a primary care setting," *Pilot and feasibility studies*, vol. 4, no. 1, p. 19, 2018.

[11] A. T. Azar and S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," *Neural Computing and Applications*, vol. 23, no. 7-8, pp. 2387–2403, 2013.

[12] L. Breiman, *Classification and regression trees*. Routledge, 2017.

[13] N. D. Vanli, M. O. Sayin, M. Mohaghegh, H. Ozkan, and S. S. Kozat, "Nonlinear regression via incremental decision trees," *Pattern Recognition*, vol. 86, pp. 1–13, 2019.

[14] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[15] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[16] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[17] M. C. Tu, D. Shin, and D. Shin, "A comparative study of medical data classification methods based on decision tree and bagging algorithms," in *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*. IEEE, 2009, pp. 183–187.

[18] W. Zhu, M. Xie, and J.-F. Xie, "A decision tree algorithm for license plate recognition based on bagging," in *Wavelet Active Media Technology and Information Processing (ICWAMTIP), 2012 International Conference on*. IEEE, 2012, pp. 136–139.

[19] P. Bonissone, J. M. Cadenas, M. C. Garrido, and R. A. Díaz-Valladares, "A fuzzy random forest," *International Journal of Approximate Reasoning*, vol. 51, no. 7, pp. 729–747, 2010.

[20] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE transactions on geoscience and remote sensing*, vol. 40, no. 10, pp. 2291–2299, 2002.

[21] M. A. Mim and K. S. Zamil, "Gis-based analysis of changing surface water in rajshahi city corporation area using support vector machine (svm), decision tree & random forest technique," *Machine Learning Research*, vol. 3, no. 2, p. 11, 2018.

[22] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.

[23] S. Michie, M. Richardson, M. Johnston, C. Abraham, J. Francis, W. Hardeman, M. P. Eccles, J. Cane, and C. E. Wood, "The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions," *Annals of behavioral medicine*, vol. 46, no. 1, pp. 81–95, 2013.

sensors

MDPI

Article

Towards a Smart Smoking Cessation App: A 1D-CNN Model Predicting Smoking Events

Maryam Abo-Tabik [1,*], Nicholas Costen [1], John Darby [1] and Yael Benn [2]

[1] Department of Computing and Mathematics, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester M15 6BH, UK; n.costen@mmu.ac.uk (N.C.); J.Darby@mmu.ac.uk (J.D.)
[2] Department of Psychology, Manchester Metropolitan University, Manchester M15 6GX, UK; Y.Benn@mmu.ac.uk
* Correspondence: maryam.a.abo-tabik@stu.mmu.ac.uk

Received: 31 December 2019; Accepted: 13 February 2020; Published: 17 February 2020

Abstract: Nicotine consumption is considered a major health problem, where many of those who wish to quit smoking relapse. The problem is that overtime smoking as behaviour is changing into a habit, in which it is connected to internal (e.g., nicotine level, craving) and external (action, time, location) triggers. Smoking cessation apps have proved their efficiency to support smoking who wish to quit smoking. However, still, these applications suffer from several drawbacks, where they are highly relying on the user to initiate the intervention by submitting the factor the causes the urge to smoke. This research describes the creation of a combined Control Theory and deep learning model that can learn the smoker's daily routine and predict smoking events. The model's structure combines a Control Theory model of smoking with a 1D-CNN classifier to adapt to individual differences between smokers and predict smoking events based on motion and geolocation values collected using a mobile device. Data were collected from 5 participants in the UK, and analysed and tested on 3 different machine learning model (SVM, Decision tree, and 1D-CNN), 1D-CNN has proved it's efficiency over the three methods with average overall accuracy 86.6%. The average MSE of forecasting the nicotine level was (0.04) in the weekdays, and (0.03) in the weekends. The model has proved its ability to predict the smoking event accurately when the participant is well engaged with the app.

Keywords: smoking cessation app; smoker's behaviour; addictive behaviour; machine learning; deep learning; CNN; control theory

1. Introduction

Smoking is considered one of the leading causes of deaths internationally. According to a recent NHS report [1], smoking caused the deaths of approximately 7900 people in England alone in 2016. The report further states that smoking is not only harmful to the smokers, but many diseases may be caused by the exposure to passive smoking, especially affecting children who are particularly vulnerable to the effects of passive smoking. This makes reducing cigarette smoking a significant public health priority. To support efficient and timely delivery of intervention for those wishing to quit smoking, it is important to be able to model the smoker's behaviour, and in order to do that, it needs to target both endogenous stressors (e.g., nicotine effect, craving, etc.)and exogenous stressors (e.g., timing, location, type of activity, etc.) that trigger the smoking events [2].

With advances in technology, new possibilities have emerged for creating efficient cessation programs, particularly through the use of mobile apps. This new technology has many advantages over traditional therapies; it can reach people wherever they are; enhance their experience by opening new channels between the therapist and the smoker; lastly, it offers the possibility to access databases

that can provide individual feedbacks on the smokers' current status [3], Several methods have been used to provide intervention using mobile apps, For example, text messages either in regular or randomized intervals, or by making the user initiate access to the intervention by reporting on indicators that may cause a potential lapse [4–6].

Investigations using self-reporting as a method have indicated that the reported predictors can provide a high degree of possibility for predicting potential lapses [4,5]. Schick et al. [6] improved this method by using Hidden Markov Models to set patterns for the timing and places in which individuals are most likely to smoke, and then use these patterns for better delivery of the support messages. This paper did not report any analytical results that are related to Hidden Markov Models, but rather focused on the positive feedback from the participants who used their mobile application.

Recent advances in computation make machine learning a perfect tool for modelling smokers' behaviour, enabling the implementation of smart mobile apps that have the ability to provide 'just in time' intervention. For example Dumortier et al. [7] used machine learning methods to evaluate the urge to smoke based on participant reporting of 41 features (e.g., alcohol consumption, mood status, hunger, location, type of working, etc.) that may trigger an urge to smoke. They compared three different machine learning algorithms (naive Bayes classifier, discriminant analysis classifier, and decision tree learning), and checked the accuracy of the classification based on a number of selected features. Results indicated that machine learning had the ability to estimate the smokers' urge rating with an accuracy of the classifications up to 86%. However, the models relied on the users reporting a large number of input features. Another study [8] also used decision tree to predict daily smoking behaviour. Here population information from the 2015 China Adult Tobacco Survey Report was used; the research modelled an equation that calculates the probability of smoking time based on gender, age and time and used statistical information from the dataset as well as some additional extracted features as input to the decision tree model. The researchers concluded that the best method of prediction is XGBoost with 84.11% accuracy.

In addition to the issues around self reporting, most existing apps for smoking cessation do not take into consideration the complexity of nicotine dependence treatment or the specific needs of the users [3]. Self-reporting as a method can be inaccurate as it is sensitive to self-biased errors based on how participants define emotional variables (e.g., withdrawal, stress, craving, alcohol use) or environmental variables (e.g., location, the presence of other smokers) [5]. Furthermore, long-term self-reporting is more likely to be affected by the 'Ostrich problem' by which people avoid monitoring their behaviour, as it may be unpleasant, tiresome, or lead to unwanted changes in behaviour [9]. Therefore, collecting time information from mobile sensors can reduce the reliance on self-reports, and increase the accuracy of just-in time intervention messages [4].

Actions (including smoking) can be seen as being motivated by the need to maintain stability over time, in the face of a changing environment. This motivation can be interrupted by internal factors, e.g., feelings such as sadness, or external factors such as nicotine level [10]. A closed-loop control model is a common instrumental technique that seeks to maintain stability. It employs a feedback principle, using the output data from the model (feedback signal) as an input to modify the model's actions, and hence maintain stability [11]. However, modelling addictive behaviour as a closed loop control model is a challenging task. It requires understanding the complexity of humans, as well as determining what elements should be counted to model the addictive behaviour. Moreover, when modelling the addictive behaviour, the goal state represents the fact that the system seeks to obtain a steady state (natural state), rather than to imply that there exists a single fixed value, as is often the case in system engineering [12,13].

Opponent process theory is claimed to be an essential method that can be used to model a person's emotional state [14]. Solomon [15] described addictive behaviour using the opponent process theory. Within this model, an addict experiences pleasure as soon as a drug is supplied, which is followed by slowly accumulated withdrawal symptoms. As such, during the initial stages of addiction, the pleasure level is high and is accompanied by a low level of withdrawal symptoms. However, as time goes

by, the withdrawal symptoms increase leading to a decrease in pleasure caused by using the drug, potentially resulting in a higher quantity of the drug being consumed [12].

Bobashev et al. [16] modelled the behaviour of smokers and employed the opponent process scheme of control theory. The model did not present any complex neurobiological process, only providing a mathematical model with a cascading feedback loop, aimed at presenting the scientific narrative of the opponent process as shown in Figure 1.

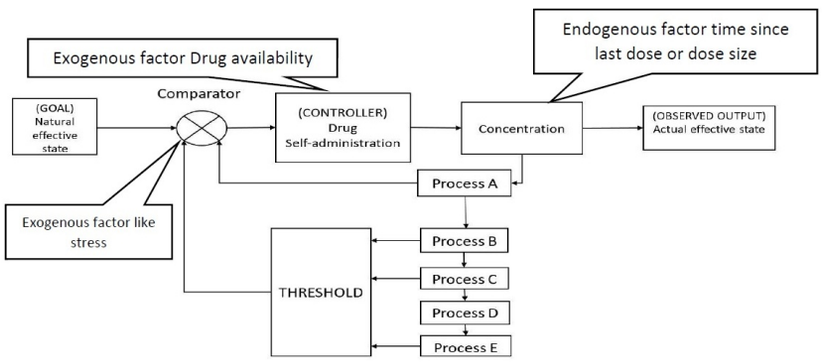


Figure 1. Control theory model of smoking re-drawn following [16].

The model equations were developed with phenomenological interpretation in mind, and no real biological process was modelled. A set of continuous functions were used, feeding into the cascading functions. The system equations involve five interlinked processes,

$$Process A : \frac{dY_1}{dt} = e^{-\alpha t} - b_1 Y_1 \tag{1}$$

$$Process B : \frac{dY_2}{dt} = a_1 Y_1 - b_2 Y_2 \tag{2}$$

$$Process C : \frac{dY_3}{dt} = a_2 Y_2 - b_3 Y_3 \tag{3}$$

$$Process D : \frac{dY_4}{dt} = a_3 Y_3 - b_4 Y_4 \tag{4}$$

$$Process E : \frac{dY_5}{dt} = a_4 Y_4 - b_5 Y_5 \tag{5}$$

where a, b and α are scaling coefficients, and all the Y_i initial values are set to zero. Each equation presents a weighted integration of the previous one, causing the processes to lengthen successively. Y_1 represents the effect of nicotine level and is modelled with a pharmacokinetic equation. Y_2 represents the toxicity level and how the body processes the drug. Y_3 is the daily smoking habit. Y_5 is a longer scaling habit, which is scaled in years (rather than minutes/hours/days). While the process Y_4 has not been interpreted, it has been used to add scaling period between Y_3 and Y_5, which results in a slow change in process Y_5. To simulate smoking behaviour, a threshold value was defined to prompt self-administration. The threshold

$$T = \frac{(\beta_3 Y_3 + \beta_5 Y_5)}{(1 + \beta_2 Y2)} \tag{6}$$

has calibration coefficients β_i, and to avoid division by zero one is added to the denominator of the equation. The threshold value is changed based on external stressors to initiate cigarette use

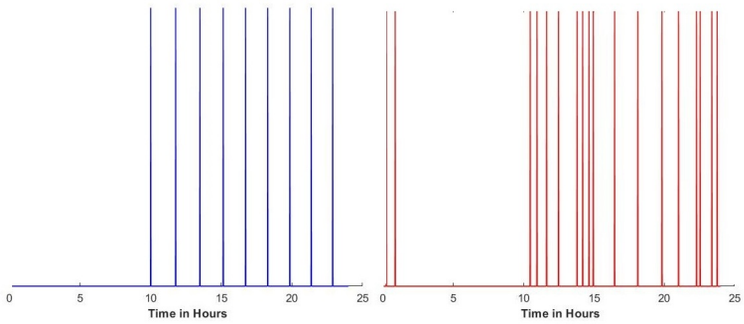
$$T = T + stress. \tag{7}$$

The research also modelled the withdrawal and craving processes; these processes begin immediately following the initial nicotine use and grow over time

$$W = \frac{d_3 Y_3 (T - Y_1)}{(Y_{0w} + Y_1)} \tag{8}$$

$$C = \frac{d_5 Y_5 (T - Y_1)}{(Y_{0c} + Y_1)} \tag{9}$$

where d_3, d_5, Y_{0w} and Y_{0c} are calibration coefficients. This control theory model was able to simulate plausible changes in smoking behaviour over time. However, the system was not able to present real-life behaviour, and could not capture individual differences between smokers' daily habits. Figure 2 shows an example of the differences between the smoking behaviour as presented using the simulated control theory model Figure 2a and real-life data collected from a participant shown in Figure 2b.



(**a**) Simulated smoking events (**b**) Real smoking events

Figure 2. Smoking frequency; each peak represents a smoking event using smoking events reported from a randomly-selected day from our collected data of one of our participants [2]. (**a**) a simulated smoking behaviour generated by the control theory model [16] , and (**b**) real smoking behaviour.

Studies show that modelling smoking behaviour is essential, as it can improve the intervention process in the way of helping smokers in their most needed time [17]. While control theory models lack in prediction but provide an explanation, on the other hand, the deep learning (DL) models provide superior prediction without explanation. In order to get better time-series data prediction, it is useful to incorporate a mechanical structure into a phenomenological statistical model [18]. Following this hypothesis, this research proposes a deep-learning model, which when combined with a control theory model of smoking, will be able to adapt to the smoker's unique behaviour and predict future smoking events. The Bobashev et al. [16] model was chosen due to its ability to capture the nicotine effect using the pharmacokinetic equation. The model can be later employed to develop a smart mobile app that will send automated interventions. Here, we describe the implementation of this control theory model of smoking that is expanded to incorporate other factors affecting smokers' smoking behaviour (e.g., geolocation and motion).

2. Classification Method: The 1D Convolutional Neural Network

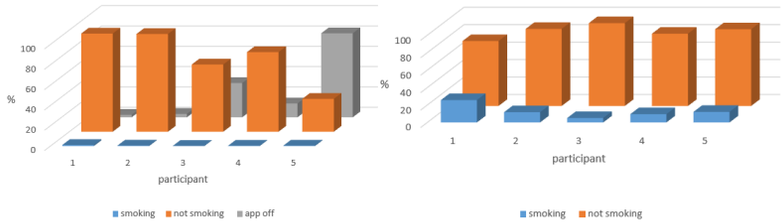
In recent years, deep learning as sub-field of machine learning (ML) has attracted great interest from the scientific community. DL refers to a deep neural network that consists of a massive web of interconnected nodes (whose depth is more than a single hidden layer). The nodes are able to perform complex, non-linear, computation on a set of input features, and give a suggested solution as an output. This new structure has been used to resolve many complex computer science problems such as image and speech recognition, with better accuracy compared to previous approaches of ML [19,20]. Convolutional neural networks (CNN) are a type of feed-forward neural network, which dates back to the 1980s. CNNs are composed of a convolution operation followed by a pooling operation [21,22]. With the increase interest in DL, CNNs have been reintroduced and used in many applications [23]. The main advantage of a CNN is its ability to be applied on parallel methods, and its high ability to learn, ensuring that all stages of the computation are appropriate for the data and for each other. To solve a problem using CNN, one should try experimenting with different variables including the number of layers, kernel size, choice of an activation function, etc. [24]. 1D-CNN performs a convolutional operation on the local region of the input data using different kernels for the individual features. Also, the size of the local region can vary for different features (this is not possible with a 2D CNN). The 1D convolutional operation in layer l,

$$y^l_{j,i} = f(\sum_m \sum_n X(j-m,n)K(m,n,i) + b) \tag{10}$$

where K is the multi-dimensional convolutional kernel, i is the kernel index, b is the bias and X and y are the input and output respectively, performs dot-products across the input [25]. In most models, a Deep CNN will use a rectified linear unit (ReLU) $f(x) = max(x, 0)$, instead of a traditional neural network (hyperbolic tangent, logistic sigmoid) activation function. ReLU is more efficient, simpler and allows non upper-bounded output values. Also, in order to improve the performance of the CNN, regularisation techniques may be used, which reduces the generalization error while preserving the training accuracy[26].

3. Data Collection and Processing

A mobile application was developed, that can collect signals from mobile sensors (e.g., movement and environment), as well as participants' self-report of smoking events. Five smokers (all taking at least 5 cigarettes per day) were recruited, and were asked to report their smoking events for two weeks. In the pre-processing stage of the data, samples for each day were unified to 1440 sample per day (one sample per minute). To do so, three types of events were registered in the dataset: smoking, not-smoking and app-off (representing gaps in the dataset due to, for example, participant's mobile phone being off). Figure 3a shows the frequency of events for each of the five participants. It is clear from the data that the classes are unbalanced, as there are far fewer smoking compared to non-smoking events. Overall, of the 1440 data samples per day less than 15 per day are smoking events, while the rest are either not smoking or app-off events. To overcome this limitation, the time periods for labelling was changed to include a 30-min window followed the smoking event rather than a 1-min window, hence reducing the ratio of smoking to non-smoking events. Furthermore, it is assumed in the model that app-off is a non-smoking event, to remain cautious. Figure 3b shows the frequency of events for each of the five participants after applying these changes.



(**a**) Percentage frequency of the three labelling categories.

(**b**) Percentage frequency after processing the data.

Figure 3. Data set Preprocessing.

The reported smoking events were then used as input to the control theory model of smoking, in order to calculate the nicotine levels and threshold value during the 13 day period (one 24 h period was dropped because it was made of two half-days, one at the start and the other at the end of the data collection period). Calculated data (e.g., nicotine level) along with collected data (e.g., light, GPS Location, activity labels etc.) were combined to form the dataset for each participant. The reported smoking events were the labels for the data set. Figure 4 illustrates the process of data collection.

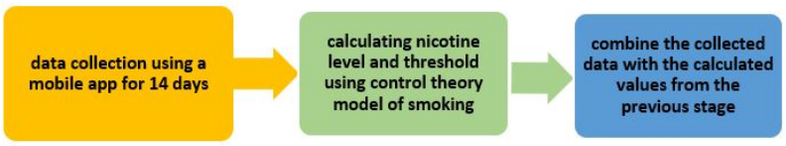
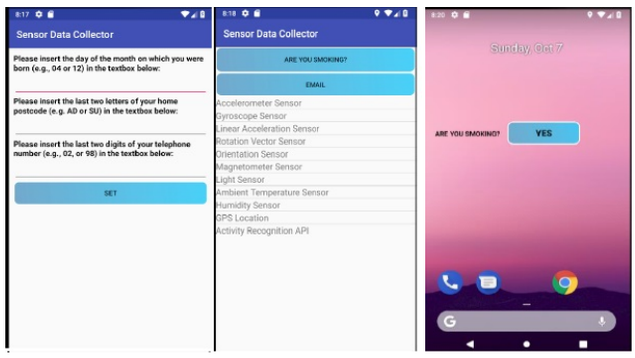


Figure 4. Overview of the study: data collection and processing steps.

3.1. Mobile App

Data collection took place using a mobile application developed for Android mobile users, using Android Studio (IDE). The main focus of the User Interface (UI) was to develop a user-friendly interface that provides no feedback to users, as so to avoid influencing their behaviour [27]. The UI was used to label smoking events, relying on participants' self-reporting. Users could report smoking events either by pressing a button on the main layout of the app, or by pressing a Widget on the home screen of the smartphone as can be seen in Figure 5.



(**a**) Participant registration screen

(**b**) Application main layout

(**c**) Smart phone home screeen

Figure 5. Mobile application User Interface (UI) [2].

The application was designed to run as a background service, which records data from the phone's sensors. This service was designed to restart itself whenever terminated (either by the OS or otherwise). This was implemented in order to overcome a new restriction forced by Android on the development of background services that run for long periods. Collected data, along with smoking events were stored on an internal SQLite database.

3.2. *Data Collection*

For this study, the participants were healthy smoking adultsover 18 years old, with a good level of English literacy. They each owned and regularly use an Android mobile phone. Smokers were defined as those smoking at least 5 cigarettes a day for at least 6 months; they all smoke traditional cigarettes. During the data collection period, the application was installed on the participant's smartphone for two weeks. No restrictions were been placed on their daily activities, and they were only asked to report their smoking events and keep the GPS on. At this stage of the research data has been collected from 5 participants (3 females: 2 male); all from the UK. The exclusion criteria were being under 18 years or over 55 years; self-reported physical or mental health issues that impact movement; not using an Android phone (e.g., using an iPhone). Although the number of participants appears small, The study by Schick et al. [6] modelled smoking behaviour using 4 participants, hence 5 participants were a sufficient number to model smoking behaviour. In addition, the ML model is trained for each participant separately, where a large volume of data was collected from each participant (approximately more than 1000 smoking events and 18720 samples each participant), making it sufficient for modelling a machine learning problem.

Data were collected from several sensors in order to identify correlations between smoking events and the sensors reading. Table 1 shows the types of collected data. The goal to use the collected data to find the association between smoking events and environmental data, in order to inform the implementation of a machine learning model that can automatically predict smoking events based on the occurrence of internal and external predictors. Following data collection, it emerged that not all sensors are available in all mobile models. Therefore the plan was modified to use only the common sensors that appear in most of the mobiles, i.e., the accelerometer and GPS values.

Table 1. The number of labels in each of the three labelling categories.

Collected Data Group Name	Description
ID	This is unique ID that Identify the user data, it is set by the user at the start of the study.
Timing value	This is time stamp DD-MMYYYY, HH:MM:SS
Motion sensors data	Accelerometer, Gyroscope, Linear acceleration, Orientation, Rotation vector.
Environmental data	Magnetic field, Light level, Ambient temperature, Relative humidity, GPS location.
Activity labels	Google activity recognition API (Still, Running, Walking, Cycling, Tilting, and Driving).
Smoking labels	This is labelled by the user.

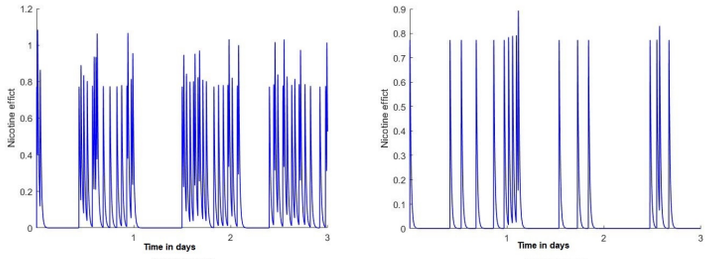
4. Approach to Model Development

To design a machine learning model for smoking behaviour, the control theory model of smoking was combined with the 1D-CNN. Initially, each part of the model was analysed separately before reaching the final model.

4.1. Control Theory Model of Smoking

While the actual nicotine level cannot be measured without lab oratory testing, that requirement does not accord with the aim of the research (creating a model that can be employed in a smart smoking cession app). The output from the control theory model of smoking is accepted as a description of the behaviour of the endogenous stressors, where the nicotine level is increased with every cigarette taken, then decreases gradually over time untill the next smoking event.

Using the reported smoking events, nicotine concentration was calculated using the control theory model of smoking [16] as shown in Figure 6. Each peak in the figure represents a smoking event, followed by a gradual decrease in the nicotine level until the next smoking event.



(**a**) Participant 1 (**b**) Participant 2

Figure 6. Examples of 3 days of smoking behaviour by two randomly selected participants, as modelled using control theory to represent nicotine levels [2].

Figure 7 shows the threshold values calculated using the control theory model. The peaks represent non-smoking periods, and the threshold value decreases as the number of cigarettes take per day increases. The control theory model also describes withdrawal and craving symptoms, as demonstrated over a 10-day period in Figure 8.

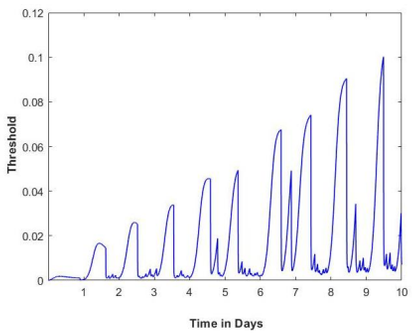
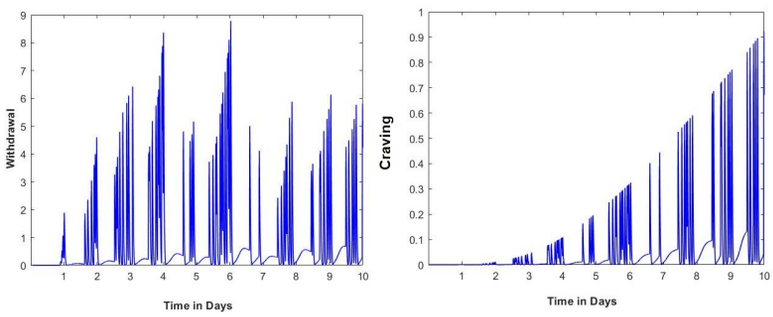


Figure 7. Example of 10 days calculated threshold value using the control theory model of smoking and collected data from one of the participants [2].



(**a**) Withdrawal value based on the control theory model of smoking.

(**b**) Craving value based on the control theory model of smoking.

Figure 8. Example of 10 days calculated withdrawal and craving values using the control theory model of smoking and collected data from one of the participants [2].

4.2. Classification of Smoker Behavioural Data

Two types of events occur in the collected dataset, which are labelled as smoking (1) and not smoking (0). In order to further verify the effectiveness of the system, three types of ML models were explored; Support Vector Machine (SVM), Decision tree (DT), and 1D-CNN. The three classifiers were implemented and tested using Python. The Scikit-learn library was used to implement SVM and DT, while 1D-CNN was implemented using the Keras neural-network library with Tensorflow-GPU in the background.

The classification methods were tested to see whether the classifier could detect the smoking events based on each smoker's motion and location factors. The data was trained and tested for each participant individually using an iterative process, where one day was held for testing while the remaining 12 were used for training. This routine was repeated separately for each of the 13 days. Six features (3 raw accelerometer values: x, y and z, and three GPS values: longitude, latitude, and altitude) were used as input to the ML models. First, the classifiers were tested using only 3 accelerometer values Table 2. Then the classifiers were tested using only the GPS values Table 3.

Table 2. The average accuracy based on only 3 GPS values.

Calculated Accuracy Category	SVM	DT	1D-CNN
Participant 1 smoking	0.01	0.40	0.01
Participant 1 not smoking	0.98	0.70	1.00
Participant 1 overall	0.73	0.62	0.74
Participant 2 smoking	0.03	0.51	0.09
Participant 2 not smoking	0.99	0.95	0.98
Participant 2 overall	0.88	0.90	0.87
Participant 3 smoking	0.02	0.08	0.00
Participant 3 not smoking	0.99	0.91	1.00
Participant 3 overall	0.94	0.95	0.95
Participant 4 smoking	0.00	0.24	0.08
Participant 4 not smoking	1.00	0.81	1.00
Participant 4 overall	0.90	0.88	0.90
Participant 5 smoking	0.00	0.25	0.21
Participant 5 not smoking	1.00	0.97	0.97
Participant 5 overall	0.88	0.97	0.88

Table 3. The average accuracy based on only 3 accelerometer values.

Calculated Accuracy Category	SVM	DT	1D-CNN
Participant 1 smoking	0.26	0.43	0.51
Participant 1 not smoking	0.74	0.75	0.83
Participant 1 overall	0.62	0.67	0.75
Participant 2 smoking	0.19	0.37	0.63
Participant 2 not smoking	0.88	0.89	0.95
Participant 2 overall	0.80	0.83	0.91
Participant 3 smoking	0.06	0.08	0.01
Participant 3 not smoking	0.95	0.94	1.00
Participant 3 overall	0.90	0.89	0.95
Participant 4 smoking	0.21	0.16	0.18
Participant 4 not smoking	0.91	0.93	0.97
Participant 4 overall	0.84	0.85	0.89
Participant 5 smoking	0.12	0.25	0.44
Participant 5 not smoking	0.95	0.95	0.94
Participant 5 overall	0.85	0.86	0.87

Table 4 shows the classification accuracy based on using motion and location factors which are 3 accelerometer (x, y and z) values and GPS (longitude, latitude, and altitude).

Table 4. The average accuracy based on all 6 featuers accelerometer and GPS values.

Calculated Accuracy Category	SVM	DT	1D-CNN
Participant 1 smoking	0.24	0.41	0.59
Participant 1 not smoking	0.79	0.77	0.79
Participant 1 overall	0.65	0.68	0.73
Participant 2 smoking	0.04	0.50	0.64
Participant 2 not smoking	0.87	0.92	0.94
Participant 2 overall	0.78	0.87	0.89
Participant 3 smoking	0.05	0.11	0.03
Participant 3 not smoking	0.96	0.93	0.99
Participant 3 overall	0.91	0.89	0.94
Participant 4 smoking	0.14	0.28	0.20
Participant 4 not smoking	0.91	0.87	0.97
Participant 4 overall	0.83	0.81	0.89
Participant 5 smoking	0.12	0.26	0.47
Participant 5 not smoking	0.95	0.95	0.94
Participant 5 overall	0.85	0.86	0.88

It can be seen from the tables that in general the performance of the 1D-CNN is consistently more accurate than the other two classifiers. The results indicate that in order to model the smoker's behaviour, the model has to be trained based on the individual behaviour for each person; and that both motion and location are important for predicting smoking events.

The 1D-CNN is implemented such that a sequence of 10 past observations is mapped as input to the model. Each input feature is passed as a 1D input to a separate model in parallel to the others, and at the end the output from these models is combined to get the classification output. The model architecture consists of 6 parallel models, each of 1D-convolutional layer with 64 filters, followed by one max-pooling layer and a flattening layer. These models combine to one dense layer with 30 nodes, the output of which is passed to a sigmoid activation function to produce the final output (see Figure 9).

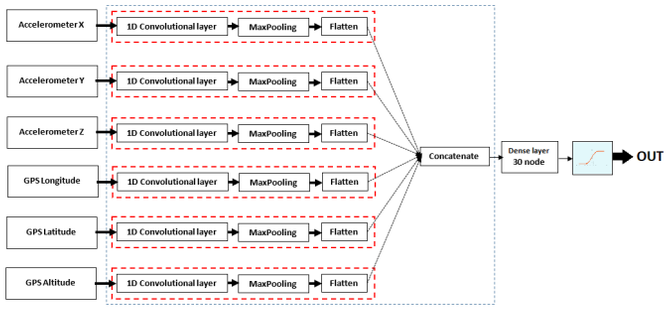
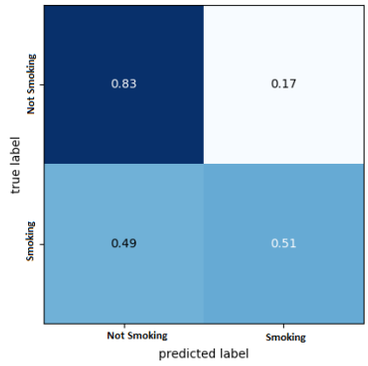


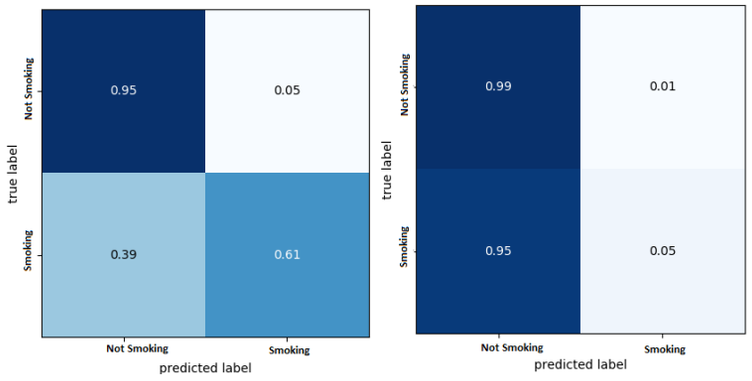
Figure 9. One-dimensional convolutional neural network (1D-CNN) model architecture.

Figure 10 shows the confusion matrix for each participant;it can be seen that the continuity of the data is an important factor that affects the prediction of smoking events. Despite this, the overall classification accuracy remains high (mean: 0.87, standard deviation: 0.080). The final model predicts nicotine levels that are much closer to the original nicotine levels for all participants, as described in the next section.



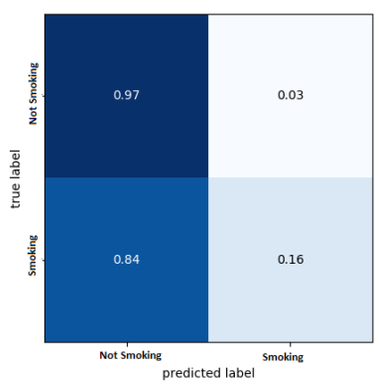
(**a**) Participant 1.

Figure 10. *Cont.*

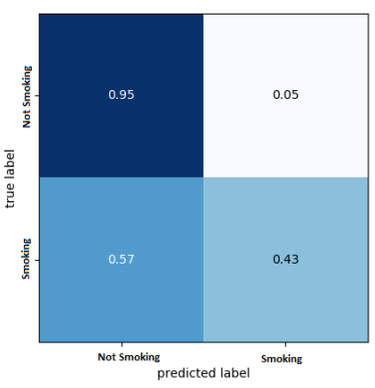


(**b**) Participant 2.

(**c**) Participant 3.



(**d**) Participant 4.



(**e**) Participant 5.

Figure 10. 1D-CNN model confusion matrix.

5. Results

After testing the three classification methods, the 1D-CNN was selected as the most suitable classifier to predict smoking events. The classifier predicts either smoking or non-smoking states, with the app off event being treated as non-smoking events. The point of the prediction was to see if the model can accurately forecast the nicotine level (rather than use the originally calculated values) using the combined control theory and 1D-CNN model, and then predict smoking events based on this predicted nicotine value. As the nicotine level is considered a changed value over time. The model uses the comparison between the nicotine level and the threshold value as the first indicator for the need to have a cigarette. The model then makes use of external stressors (accelerometer and GPS) as input to the DL model in order to make the final decision regarding the likelihood of a smoking event. The importance of using the 1D-CNN model as part of the control theory model is to ensure the capture of the endogenous factors which affect the smoker's behaviour as presented by nicotine level inside the smoker's body. This approach should ensure that no intervention messages are sent

before the nicotine level as derived from Equation (1) decreases to a level that is below the threshold as derived from Equation (6).

The resultant combined model of DL and control theory is shown in Figure 11. Six features were used as input to the DL model (three raw accelerometer values: x, y and z and three GPS values: longitude, latitude, and altitude). Since all these values and their combinations are personalised for each participant, training needs to take place for each participant. Testing the data iteratively enabled us to compare the prediction level for different days of the week. Since the output of the model is forecasting the nicotine effect value over time, Mean Square Error is used as the error criterion to measure the performance of the model. This evaluation matrix has been previously used to evaluate time-series data [28]. The results of the Mean Square Error (MSE), Root Mean Square Error (RMSE), and Normalized Root Mean Square Error (NRMSE), represent the accuracy of predicted nicotine levels during week days (Table 5) and weekends (Table 6). In general, the model has the same performance throughout the week.

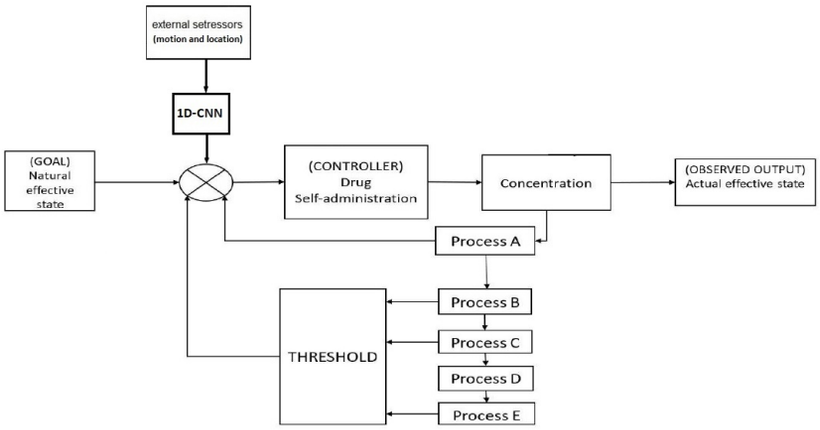


Figure 11. Smoking behaviour model utilizing machine learning. Data are collected and processed using the steps described in Figure 4. The 2 predictors are used as input to the 1D-CNN model. A classification value of 1 represents a potential smoking event. This value is passed to the CONTROLLER, simulating the taking of a cigarette, and re-initializing the parameters of the control model to zero.

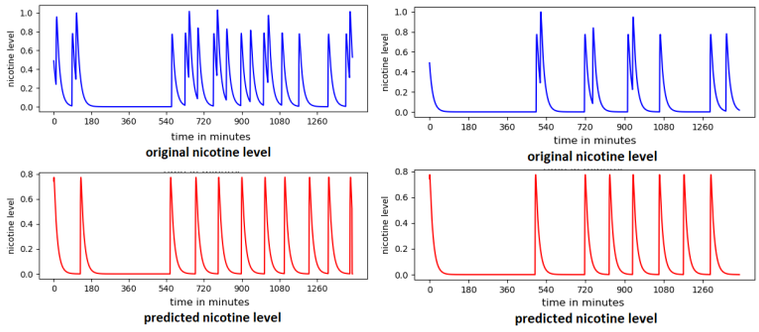
Table 5. The overall error rate of the proposed model during week days days.

	MSE	RMSE	NRMSE
Participant 1 smoking	0.07	0.26	0.23
Participant 2 smoking	0.05	0.22	0.20
Participant 3 smoking	0.02	0.14	0.15
Participant 4 smoking	0.03	0.18	0.21
Participant 5 smoking	0.03	0.15	0.17

Table 6. The overall error rate of the proposed model over the weekends.

	MSE	RMSE	NRMSE
Participant 1 smoking	0.07	0.27	0.24
Participant 2 smoking	0.03	0.16	0.17
Participant 3 smoking	0.01	0.10	0.11
Participant 4 smoking	0.03	0.16	0.20
Participant 5 smoking	0.01	0.12	0.15

Figure 12 shows the predicted nicotine level from a randomly selected day for two participants. All 6 predictors were used as input to the system. The nicotine level was predicted during the closed-loop process; no pre-calculated data was used.

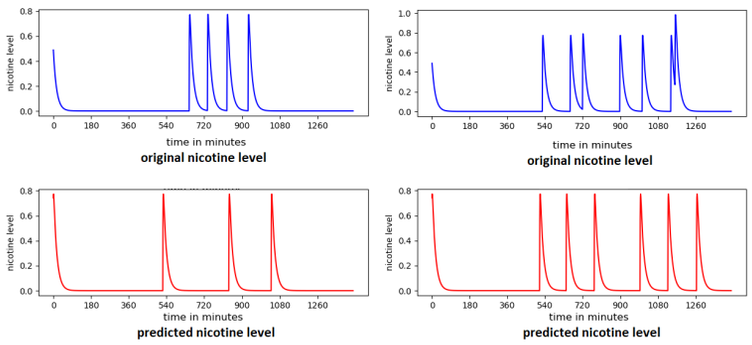


(**a**) Randomly selected one day from participant 1.

(**b**) Randomly selected one day from participant 2.

Figure 12. Predicted nicotine level from two participants.

Although some smoking events were missed, the model in general reliably predicts the smoking behaviour of each of the participants. While accuracy of prediction of nicotine level is negatively affected by missed samples in the data set overall accuracy remains high. Figure 13 shows the predicted smoking events for a randomly selected day with a high level of missed samples.



(**a**) Randomly selected one day from participant 3.

(**b**) Randomly selected one day from participant 4.

Figure 13. Predicted nicotine level for participants with a high presence of app-off values.

In some cases the participant was cooperative in reporting smoking events in all days accept for one day. The model predicted several smoking events for that day, and we cannot be sure whether these are unreported smoking events or false alarms (Figure 14).

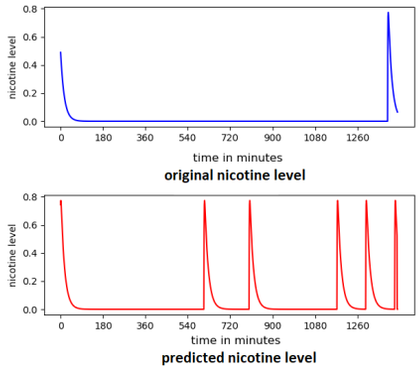
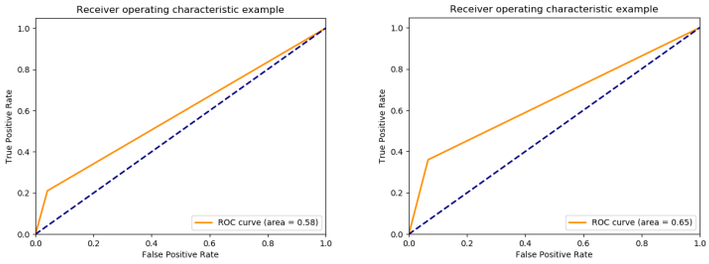
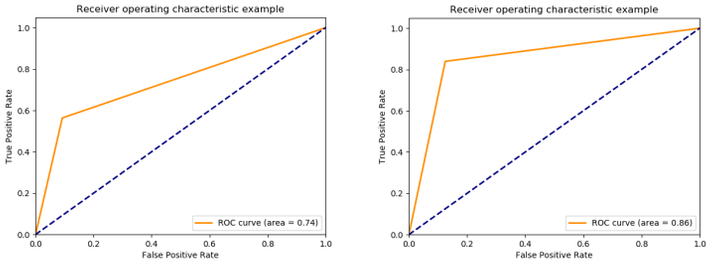


Figure 14. Predicted nicotine level for one day from participant 3 with no reported smoking events.

Overall, the model can predict smoking events with 0.2 accuracy in a 15 min window, 0.3 accuracy in a 30 min window, 0.5 in a 1-h window, and 0.8 in a 2-h window. Figure 15 displaysthe ROC curve for all dataset for 15 min, 30 min, 1-h window, and 2-h window. As far as we know, there are relatively few studies which explore the possibility of using machine learning to classify the factors that lead to smoking events, and all the previous research [7,8] rely on self-reporting and surveys, which make it hard to compare with our research since they use different experiment settings and different inputs. Even though this research has accomplished a better overall accuracy equals to 86.6% without relying on self-reporting of predictors.



(**a**) ROC curve with 15 min window. (**b**) ROC curve with 30 min window.



(**c**) ROC curve with 1-h window. (**d**) ROC curve with 2-h window.

Figure 15. The deep learning model of smokers behaviour ROC curve.

6. Conclusions

In conclusion, DL was successfully applied to model smokers' behaviour. The model design combines 1D-CNN with a control theory model of smoking, and the results are promising. Two predictors were used as input to 1D-CNN (motion and geolocation) to predict smoking events. This design was able to adapt to the behaviour of individual smokers, with an average of 87% overall accuracy. This is a preliminary model, with potentials of improvement in the future. It is expected that the accuracy of the system in predicting smoking events will be increased by improving the DL model by adding layers, and by taking advantage of other information such as the indoor smoking ban in the UK. It may also be possible to construct a combined model of individuals' behaviour, using additional external data such as the addresses of the smoker's work and home, as well as public information on the location of businesses such as bars and tobacco shops, which are likely to be associated with smoking. These additions to the model are currently under consideration. This model is aimed to be used by smoking cessation app; the app can be integrated in the future into smoking cessation treatments. While all previous work in this area, even if they used ML models, they continue to rely on self-reporting of the predictors. Using DL opens the door to the possibility of automatic predicting the smoker's behaviour, and that in turn allows sending automated innovations based on the individuals' behaviour.

Author Contributions: Conceptualization, M.A.-T., N.C., J.D. and Y.B.; software, M.A.-T. and J.D.; validation, M.A.-T.; formal analysis, M.A.-T. data curation, M.A.-T., Y.B. and N.C..; writing–original draft preparation, M.A.-T.; writing–review and editing, Y.B. and N.C.; visualization, M.A.-T. and J.D.; supervision, N.C. and Y.B.; project administration, M.A.-T., N.C. and Y.B.; funding acquisition, M.A.-T., N.C. and Y.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Acknowledgments: We thank the Manchester Metropolitan University intern students Olivia Faye Salmon and Alex Barrett-Chapman, employed under the funding program from the faculty of Health, Psychology and Social Care UKRIS and the Faculty of Science pump funding, who assisted with the data collection for this study, and also the participants who gave up their time and phone battery power for this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. NHS. *Adult Smoking Habits in the UK: 2016*; Office of national statistics: London, UK, 2018.
2. Abo-Tabik, M.; Costen, N.; Darby, J.; Benn, Y. Decision Tree Model of Smoking Behaviour. In Proceedings of the 2019 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation. IEEE, 2019, Leicester, UK, 19–23 August 2019; pp. 1746–1753.
3. McClure, J.B.; Hartzler, A.L.; Catz, S.L. Design considerations for smoking cessation apps: Feedback from nicotine dependence treatment providers and smokers. *JMIR mHealth uHealth* **2016**, *4*, e17. [CrossRef] [PubMed]
4. Naughton, F.; Hopewell, S.; Lathia, N.; Schalbroeck, R.; Brown, C.; Mascolo, C.; McEwen, A.; Sutton, S. A context-sensing mobile phone app (Q sense) for smoking cessation: A mixed-methods study. *JMIR mHealth uHealth* **2016**, *4*, e106. [CrossRef] [PubMed]
5. Businelle, M.S.; Ma, P.; Kendzor, D.E.; Frank, S.G.; Wetter, D.W.; Vidrine, D.J. Using intensive longitudinal data collected via mobile phone to detect imminent lapse in smokers undergoing a scheduled quit attempt. *J. Med. Internet Res.* **2016**, *18*, e275. [CrossRef] [PubMed]
6. Schick, R.S.; Kelsey, T.W.; Marston, J.; Samson, K.; Humphris, G.W. MapMySmoke: Feasibility of a new quit cigarette smoking mobile phone application using integrated geo-positioning technology, and motivational messaging within a primary care setting. *Pilot Feasibility Stud.* **2018**, *4*, 19. [CrossRef] [PubMed]
7. Dumortier, A.; Beckjord, E.; Shiffman, S.; Sejdić, E. Classifying smoking urges via machine learning. *Comput. Methods Programs Biomed.* **2016**, *137*, 203–213. [CrossRef] [PubMed]
8. Zhang, Y.; Liu, J.; Zhang, Z.; Huang, J. Prediction of Daily Smoking Behavior Based on Decision Tree Machine Learning Algorithm. In Proceedings of the 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 12–14 July 2019; pp. 330–333.
9. Webb, T.L.; Chang, B.P.; Benn, Y. 'The Ostrich Problem': Motivated avoidance or rejection of information about goal progress. *Soc. Personal. Psychol. Compass* **2013**, *7*, 794–807. [CrossRef]
10. Fibla, M.S.; Bernardet, U.; Verschure, P.F. Allostatic control for robot behaviour regulation: An extension to path planning. In Proceedings of the Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, Taipei, Taiwan, 18–22 October 2010; pp. 1935–1942.
11. Hughes, J.M. *Real World Instrumentation with Python: Automated Data Acquisition and Control Systems*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2010.
12. Gutkin, B.; Ahmed, S.H. *Computational Neuroscience of Drug Addiction*; Springer Science & Business Media: Berlin, Germany, 2011.
13. Ahmed, S.H.; Bobashev, G.; Gutkin, B.S. The simulation of addiction: Pharmacological and neurocomputational models of drug self-administration. *Drug Alcohol Depend.* **2007**, *90*, 304–311. [PubMed]
14. Metin, S.; Sengor, N.S. From occasional choices to inevitable musts: A computational model of nicotine addiction. *Comput. Intell. Neurosci.* **2012**, *2012*, 18. [CrossRef] [PubMed]
15. Solomon, R.L. The opponent-process theory of acquired motivation: The costs of pleasure and the benefits of pain. *Am. Psychol.* **1980**, *35*, 691. [CrossRef] [PubMed]
16. Bobashev, G.; Holloway, J.; Solano, E.; Gutkin, B. A Control Theory Model of Smoking. *Methods Rep. (RTI Press)* **2017**, *2017*. [CrossRef]
17. Timms, K.P.; Rivera, D.E.; Collins, L.M.; Piper, M.E. Control systems engineering for understanding and optimizing smoking cessation interventions. In Proceedings of the 2013 American Control Conference, Washington, DC, USA, 17–19 June 2013; pp. 1964–1969.
18. Bobashev, G.V.; Ellner, S.P.; Nychka, D.W.; Grenfell, B.T. Reconstructing susceptible and recruitment dynamics from measles epidemic data. *Math. Population Stud.* **2000**, *8*, 1–29. [CrossRef]

19. Popa, D.; Pop, F.; Serbanescu, C.; Castiglione, A. Deep learning model for home automation and energy reduction in a smart home environment platform. *Neural Comput. Appl.* **2019**, *31*, 1317–1337. [CrossRef]
20. Zahid, M.; Ahmed, F.; Javaid, N.; Abbasi, R.A.; Kazmi, Z.; Syeda, H.; Javaid, A.; Bilal, M.; Akbar, M.; Ilahi, M. Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids. *Electronics* **2019**, *8*, 122. [CrossRef]
21. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
22. Kalchbrenner, N.; Grefenstette, E.; Blunsom, P. A Convolutional Neural Network for Modelling Sentences. Available online: https://arxiv.org/abs/1404.2188 (15 February 2020)
23. Nguyen, T.H.; Grishman, R. Relation extraction: Perspective from convolutional neural networks. In Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, Denver, Colorado, 31 May–5 June 2015; pp. 39–48.
24. Sorokin, M.; Zhdanov, A.; Zhdanov, D. Recovery of Optical Parameters of a Scene Using Fully-Convolutional Neural Networks. Available online: http://ceur-ws.org/Vol-2344/short1.pdf (15 February 2020)
25. Deka, B. *Pattern Recognition and Machine Intelligence: 8th International Conference, PReMI 2019, Tezpur, India, December 17–20, 2019, Proceedings, Part I*; Springer Nature: Berlin, Germany, 2019.
26. Camps, J.; Sama, A.; Martin, M.; Rodriguez-Martin, D.; Perez-Lopez, C.; Arostegui, J.M.M.; Cabestany, J.; Catala, A.; Alcaine, S.; Mestre, B.; et al. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowl. Based Syst.* **2018**, *139*, 119–131. [CrossRef]
27. Michie, S.; Richardson, M.; Johnston, M.; Abraham, C.; Francis, J.; Hardeman, W.; Eccles, M.P.; Cane, J.; Wood, C.E. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Ann. Behav. Med.* **2013**, *46*, 81–95. [CrossRef] [PubMed]
28. Assaad, M.; Boné, R.; Cardot, H. A new boosting algorithm for improved time-series forecasting with recurrent neural networks. *Inf. Fusion* **2008**, *9*, 41–55. [CrossRef]