Computer Methods and Programs in Biomedicine 217 (2022) 106638

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Knowledge-driven feature engineering to detect multiple symptoms using ambulatory blood pressure monitoring data

Zaffar Haider Janjua[a,*], David Kerins[b], Brendan O'Flynn[a], Salvatore Tedesco[a]

[a] Tyndall National Institute, Cork, Ireland
[b] University College Cork, Cork, Ireland

A R T I C L E I N F O

Article history:
Received 4 May 2021
Revised 14 November 2021
Accepted 14 January 2022

Keywords:
Multi-label classification
Knowledge-driven feature extraction
Hypertension
Ambulatory blood pressure monitoring
Feature engineering

A B S T R A C T

Background: Hypertension is a major health concern across the globe and needs to be properly diagnosed to so it can be treated and to mitigate for this critical health condition. In this context, ambulatory blood pressure monitoring is essential to provide for a proper diagnosis of hypertension, which may not be possible otherwise due to the white coat effect or masked hypertension. In this paper, the objective is to develop a model which incorporates expert's knowledge in the feature engineering process so as to accurately predict multiple medical conditions. As a case study, we have considered multiple symptoms related to hypertension and used an ambulatory blood pressure monitoring method to continuously acquire hypertension relevant data from a patient. The goal is to train a model with a minimum set of the most effective knowledge-driven features which are useful to detect multiple symptoms simultaneously using multi-class classification techniques.

Method: Artificial intelligence-based blood pressure monitoring techniques introduce a new dimension in the diagnosis of hypertension by enabling a continuous (24hours) analysis of systolic and diastolic blood pressure levels. In this work, we present a model that entails a knowledge-driven feature engineering method and implemented an ambulatory blood pressure monitoring system to diagnose multiple cardiac parameters and associated conditions simultaneously these include morning surge, circadian rhythm, and pulse pressure. The knowledge-driven features are extracted to improve the interpretability of the classification model and machine learning techniques (Random Forest, Naive Bayes, and KNN) were applied in a multi-label classification setup using RAkEL to classify multiple conditions simultaneously.

Results: The results obtained ($F1 = 0.918$) show that the Random forest technique has performed well for multilabel classification using knowledge-driven features. Our technique has also reduced the complexity of the model by reducing the number of features required to train a machine learning model.

Conclusion: Considering these results, we conclude that knowledge-driven feature engineering enhances the learning process by reducing the number of features given as input to the machine learning algorithm. The proposed feature engineering method considers expert's knowledge to develop better diagnosis models which are free from misleading data-driven noisy features in some situations. It is a white-box approach in which clinicians can under stand the importance of a feature while looking at its value.

1. Introduction

Cardiovascular diseases (CVDs) are a major cause of death across the globe. An estimated 17.9 million deaths are observed annually across the world due to CVDs [1,2]. Hypertension is the most common risk factor for CVD and it is considered to be a silent killer, because it can affect a person without them having any visible symptoms, making it an extremely dangerous clinical condition. The International Society of Hypertension (ISH) has published its guidelines to control blood pressure through multiple methods including food intake, increasing levels of exercise and, sleep, and maintaining an overall healthy life style [3]. The timely detection of elevated blood pressure is important to avoid health complica-

* Corresponding authors.
E-mail addresses: zaffar.janjua@tyndall.ie (Z.H. Janjua), D.Kerins@ucc.ie (D. Kerins), brendan.oflynn@tyndall.ie (B. O'Flynn), salvatore.tedesco@tyndall.ie (S. Tedesco).

https://doi.org/10.1016/j.cmpb.2022.106638
0169-2607/

tions and life-threatening conditions. Blood pressure levels, when measured in a clinic give clinicians a snapshot of blood pressure at the instant of measurement and this varies across the duration of the measurement. As such, measurements taken over a short time window can be misleading and do not give clinicians a clear picture regarding the true overall level of blood pressure, with potential overestimation, often called "white-coat effect" or underestimation of blood pressure known as "masked hypertension". Due to this reason, ambulatory blood pressure monitoring (ABPM) is considered a safe and reliable method for blood pressure measurement and is the recommended procedure to measure blood pressure variations as the appropriate diagnosis of the disorder [4]. The technique uses an inflatable blood pressure cuff monitoring device which records blood pressure every 10 – 15 minutes over a period of 24 hours using oscillometric method. In this way, clinicians acquire an overall picture of the patient's condition including diurnal variations identified through analysis of daytime and night-time measurements.

The daily continuous monitoring of blood pressure is essential because of many of the symptoms of CVDs are related to irregular blood pressure variations such as *morning surge, pulse pressure, and circadian rhythm*. These variations can be simultaneously present in a patient and, thus, their timely detection could help clinicians to diagnose and treat to reduce or prevent the onset of complicated medical conditions. With ABPM readings, clinicians have a better understanding of the patient's health and, indeed several studies have been conducted to use ABPM measurements for effectively diagnosing CVDs [5–7].

In this paper, an Artificial Intelligence-based technique is presented to achieve two objectives: a) it utilizes the knowledge of an appropriate domain experts to extract meaningful features from datasets which significantly contribute to the detection of a clinical condition due to blood pressure variations, b) it detects multiple symptoms of blood pressure (multi-label classification) simultaneously using ABPM data while applying the acquired experts' knowledge. Our proposed model addresses these two important problems involved in clinical data analysis, which are useful for the accurate prediction of clinical conditions.

The use of domain expert's knowledge-based feature engineering has a particular significance in the medical domain. In the medical domain, feature engineering is a sensitive task and requires a particular care when compared to other fields such as computer vision, and natural language processing. Data-driven feature engineering tools (e.g. Cognito, tsfresh) could be cost-effective, however, the accuracy and transparency of the predictive model could potentially be compromised when using data-driven feature engineering as some features are not medically significant or necessarily relevant in the prediction of a medical condition [8]. Knowledge-guided features could be more comprehensive and improve the performance of the computational model as knowledge-driven methods includes clinical expert's knowledge related to the medical condition as part of the overall process [9].

The second objective of this work is to study the influence of a domain experts' knowledge on the predictive model for multi-symptoms classification. It is observed that human health disorders are often diagnosed through the combination of multiple symptoms and the collective consideration of these symptoms can help clinicians to identify a particular disorder. Machine learning models could help clinicians to reliably foresee the existence of multiple symptoms simultaneously using the health-related data for the given disorder.

To the best of our knowledge, such knowledge-driven feature engineering and multi-label classification has not been applied together on clinical data to simultaneously detect multiple ABPM symptoms. The rest of the paper is divided into the following sections. Section 4 describes research which has already been conducted in the area of knowledge-driven feature engineering and multi-label classification. Section 2 presents a framework which combines knowledge-driven feature engineering and multi-label classification to detect multiple symptoms related to ABPM. Section 3 describes experiments we have conducted using the ABPM dataset explained in Subsection 3.1. The results obtained are presented in Section 3.3 and discussed in Section 4. Conclusions are finally drawn in Section 5.

2. Methods

The main objective of this research is to investigate the impact of incorporating clinical experts' knowledge during the feature engineering process. In this perspective, the novel contribution of our work is the extraction of knowledge-driven features from ambulatory blood pressure data and subsequently using these features for multiple symptoms classification. Formally, we defined it as a multi-label classification problem, where each feature instance carries multiple labels. Our case study is based on ABPM data, which was collected for 24 hours, recording systolic and diastolic blood pressure measurements. The primitive ABPM data reflect physical variations in systolic and diastolic blood pressure, without incorporating any experts' knowledge explaining the state of collected figures. The primitive ABPM data is defined by time-domain features showing blood pressure measurements carried out during different time intervals (e.g. daytime and night-time systolic and diastolic blood pressure recordings). The ABPM data can be defined as a time-domain feature set F, where $F = \{f_1, f_2, f_3, \ldots, f_n\}$. Each feature f_i is based on ABPM measurements acquired from a group of patients. The dataset adopted in this study is further discussed in Section 3.1. Knowledge-driven features are extracted after applying experts' knowledge D on the feature set F, therefore, creating a new knowledge-driven feature set F_D. According to the problem formulation, the knowledge-driven feature set F_D is associated with multiple labels L, where $L = \{l_1, l_2, \ldots, l_n\}$. A machine learning algorithm tries to find a hidden relationship between instances of feature set F_D and the associated label set L by using a training set. To solve this problem, we explored multi-label classification techniques and considered an improved version of Label powerset [10] method which is called RAkEL (RAndom k labELsets) [11].

The proposed framework illustrated in Fig. 1, which shows the conceptual model that was developed to include domain knowledge in the design process. In the medical domain, there are multiple knowledge sources such as clinicians, web-based clinical databases, books, and journals. For the sake of simplicity, in this work, we relied on a *manual knowledge extraction* method. The manual knowledge extraction was carried out by a *knowledge engineer* who acquired knowledge through multiple methods such as interviews, web searches, and reading medical journals. The extracted knowledge is translated into *first-order logic rules*. We have considered blood pressure related knowledge in this work. The blood pressure varies over 24 hours for a variety of reasons. For example, overnight blood pressure is normally lower as compared to daytime measurements, when the person is carrying out activities, this is known as "nocturnal dipping". The domain knowledge extracted from various sources guides the feature extraction process, which could improve the multi-symptom classification process.

Taking into account the knowledge related to blood pressure, we manually formulated several *IF-THEN decision rules*. These rules were stored in a rule-base and applied to the primitive time-domain ambulatory blood pressure data to obtainn new knowledge-driven features that could have a better relevance to the multiple labels for the classification learning model. For example, we defined some rules while considering the knowledge mentioned in Table 1. The table shows knowledge about blood pressure range divided into 5 categories: optimal, normal, high normal,

Table 1
Definition of hypertension grades [12].

Category	Systolic BP range (mmHg)	Condition	Diastolic BP range (mmHg)
Optimal	< 120	*and*	< 80
Normal	120 – 129	*and/or*	80 – 84
High normal	130 – 139	*and/or*	85 – 89
Grade 1 hypertension	140 – 159	*and/or*	90 – 99
Grade 2 hypertension	160 – 179	*and/or*	100 – 109
Grade 3 hypertension	≥ 180	*and/or*	≥ 110
Isolated systolic hypertension	≥ 140	*and*	< 90

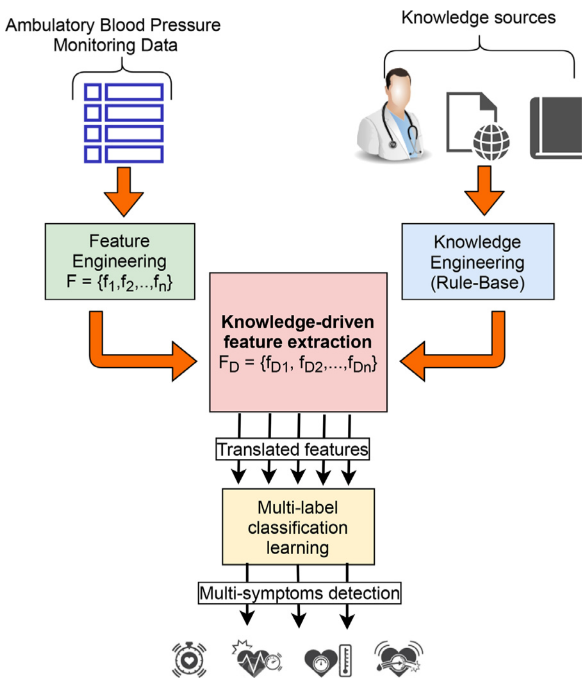

Fig. 1. Proposed framework for knowledge-driven multi-symptoms detection.

grade 1, grade 2, and grade 3. This knowledge was acquired from a document published jointly by the *European Society of Hypertension* and the *European Society of Cardiology* [12]. As mentioned earlier, the rule-base contains many many similar knowledge-driven rules, and the rules represented in sets 1,2, and 3 is an excerpt taken from the whole rule-base.

More explicitly, the rules defined in set 1 are related to average ambulatory blood pressure measurements. If the average of 24 hours systolic ambulatory blood pressure reading is less than 120 mm Hg and diastolic is less than 80 mm Hg then it is considered an optimal blood pressure. Similarly, other rules in set 1 are defined according to the knowledge mentioned in Table 1.

$$\begin{aligned}
&if\ (systolic < 120 \wedge diastolic < 80) \Rightarrow Optimal\\
&if\ (systolic \geq 120 \wedge systolic \leq 129) \Rightarrow Normal\\
&if\ (systolic \geq 130 \wedge systolic \leq 139) \Rightarrow High_Normal\\
&if\ (systolic \geq 140 \wedge systolic \leq 159) \Rightarrow Grade1\\
&if\ (systolic \geq 160 \wedge systolic \leq 179) \Rightarrow Grade2\\
&if\ (systolic \geq 180) \Rightarrow Grade3
\end{aligned} \tag{1}$$

In the case of set 2, knowledge related to *blood pressure load* (*bp_load_day*) along with *maximum systolic blood pressure* (*max_systolic*) value is taken into account. Blood pressure load is defined as the percentage of systolic blood pressure values measured above than 140 mm Hg in 24 hours [13]. *max_systolic* is the maximum value of the systolic blood pressure measurements during day and night-time. The combination of these two features give us information regarding the critical situation of the subject. In Set 2, the categories are defined by the day time blood pressure measurements. Similarly, the Set 3 represents the knowledge-driven rule for the night time blood pressure. Blood pressure load γ is tested for different values, and finally selected $\gamma = 15$ for day time and $\gamma = 5$ for night time measurements, to obtain optimized classification results.

$$\begin{aligned}
&if\ (max_systolic < 140) \Rightarrow Normal\\
&if\ (bp_load_day < \gamma = 15)\wedge\\
&\quad (max_systolic \geq 140 \wedge max_systolic \leq 159) \Rightarrow Grade1\\
&if\ (bp_load_day < \gamma = 15)\wedge\\
&\quad (max_Systolic \geq 160 \wedge max_systolic \leq 179) \Rightarrow Grade2\\
&if\ (bp_load_day < \gamma = 15)\wedge\\
&\quad (max_systolic \geq 180) \Rightarrow Grade3
\end{aligned} \tag{2}$$

$$\begin{aligned}
&if\ (max_Systolic < 140) \Rightarrow Normal\\
&if\ (bp_load_night < \gamma = 5)\wedge\\
&\quad (max_systolic \geq 140 \wedge max_systolic \leq 159) \Rightarrow Grade1\\
&if\ (bp_load_night < \gamma = 5)\wedge\\
&\quad (max_systolic \geq 160 \wedge max_systolic \leq 179) \Rightarrow Grade2\\
&if\ (bp_load_night < \gamma = 5)\wedge\\
&\quad (max_systolic \geq 180) \Rightarrow Grade3
\end{aligned} \tag{3}$$

Mathematically, let $D_t = \{(F_{D_i}, L_i) | i = 1, \ldots, N\}$ is the given labelled data. we are interested in finding the conditional probability of labelset L given features F_D based on domain knowledge D, i.e. $p(L|F_D)$. The rules defined in rule-base are applied to the raw ambulatory blood pressure feature set F to extract new knowledge-driven features F_D. For example, if the measured blood pressure is 120 mm Hg (systolic) and 80 mm Hg (diastolic) then both features are translated into a new knowledge-driven feature $F_{D_i} \in F_D$, which takes an optimal value defined according to clinician's knowledge mentioned in Table 1. To detect multiple symptoms represented as labels in L, we have conducted experiments using different supervised machine learning techniques, which learn prediction classes using an annotated dataset. First, we have considered a Naïve Bayes classifier within a multi-label classification setup. The Naïve Bayes approach models the mathematical problem and calculates the *posterior probability* $p(L|F_D)$ based on knowledge-driven *prior* and *likelihood* probability. Due to the discrete nature of features F_D, we have used a *Multinomial* implementation of Naïve Bayes classifier [14]. The states of these features are optimal, normal, high normal, grade 1, grade 2, and grade 3. These state values are given based on the clinical definition of blood pressure values.

To generalize our approach, we have also used other classifiers to predict the multiple labels. We have considered a *Random Forest classifier*, based on the decision tree algorithm in which the strongest estimators are selected through a bootstrap and bagging method, and *K nearest neighbour (KNN)* algorithm, based on distance estimation between k training examples in feature space. Table 2 lists the knowledge-driven features extracted by using the proposed approach. As already explained, these features are derived from statistical ambulatory blood pressure features listed in Table 3. The new knowledge-driven features mainly state the condition of a person as identified by a clinical expert after observing the clinical ambulatory blood pressure data such as if the person

Table 2
Knowledge-driven features extracted from ambulatory blood pressure dataset. The discrete states of features are optimal, normal, high normal, grade 1, grade 2, grade 3.

No.	Feature	Description
1	*BPS_Average*	Average of 24 hours systolic blood pressure
2	*BPD_Average*	Average of 24 hours diastolic blood pressure
3	*BPS_Day*	Average of day time systolic blood pressure
4	*BPD_Day*	Average of day time diastolic blood pressure
5	*BPS_Night*	Average of night time systolic blood pressure
6	*BPD_Night*	Average of night time diastolic blood pressure
7	*BPS_load_Day*	Systolic blood pressure load day time
8	*BPS_load_Night*	Systolic blood pressure load night time

has normal blood pressure or it lies in a critically ill category. The proposed model uses these new knowledge-based features to simultaneously detect multiple conditions: circadian rhythm, pulse pressure, and morning surge.

3. Results

In this section, we present the experimental setup alongwith acquired results to validate the performance of the proposed model. Section 3.1 explains the dataset used to perform the experiments and Section 3.2 explains all the tools and techniques used to perform the experiments.

3.1. Dataset

The dataset D_t [15] used for experiments is a multi-label mixed gender ambulatory blood pressure dataset having multiple labels for each instance. We have considered a subset of labels to validate our proposed technique related to cardiovascular disorder and discarded unrelated labels. The selected labels were circadian rhythm, pulse pressure, and morning surge. All labels have two classes: true and false. Circadian rhythm has 97 true cases and 173 false cases; pulse pressure has 232 true cases and 38 false cases; morning surge has 37 true cases and 233 false cases. Circadian rhythm relates to the absence of nocturnal blood pressure decrease. Normally, during sleep blood pressure is $10-20\%$ less than the day time blood pressure [16]. However, in some cases, this decrease in blood pressure is absent $(0-10\%)$. This absence may increase the risk of cardiovascular disorders. The second label, pulse pressure, is the difference between systolic blood pressure and diastolic blood pressure [17]. The difference indicates cardiovascular problems in elderly persons. If this difference is more than 50 mm Hg then it is an alarming situation for hypertensive patients older than 50 years. The third label, morning surge, is a measurement of the rise in blood pressure from its lowest value during sleep to the first 2 hours after wakeup [18]. With these labels, we define our label set as $L = \{l_1, l_2, l_3\}$. The remainder of the dataset is described by 40 attributes which are mainly based on ambulatory blood pressure readings such as average systolic and diastolic blood pressure readings over 24 hours, day time systolic and diastolic measurements, night time systolic and diastolic blood pressure measurements, day and night time blood pressure systolic and diastolic load values, maximum and minimum systolic and diastolic blood pressure values. Each row in the dataset represents one person with a numerical value of the attribute.

3.2. Implementation

We have implemented the framework shown in Fig. 1 in Python by primarily using two libraries: Scikit-learn [19] and Scikit-multilearn [20]. We have used scikit-learn functions for machine learning techniques: MultinomialNB (Naïve Bayes) classifier, RandomForestClassifier, and KNeighborsClassifier. RAkEL (RAndom k labELsets) [11] technique is used to implement the multi-label environment. RAkEL is an extension of the well known multi-label classification powerset technique. In the case of powerset, each distinct set of label is considered as a class and the machine learning model tries to learn one labelset. In this way, the multi-label classification is actually transformed into single label classification. However, the complexity of powerset technique increases as the number of distinct labelset patterns increases. To solve this problem, RAkEL divides the initial set of labels into smaller random k subsets and the machine learning technique try to find association between features and label subsets.

3.3. Experimental results

The results obtained from the experiments are shown in Table 4. We evaluated the performance of each technique using Precision, Recall, F1-score, Hamming loss, Jaccard score. Mathematically,Precision, Recall, and F1-score are defined by the following equations.

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{5}$$

$$F1 = 2 \cdot \frac{(precision \cdot recall)}{(precision + recall)} \tag{6}$$

Table 4 shows the score obtained for different machine learning techniques along with the aacquired parameter settings, where required. Fig. 2 shows the graphical representation of scores mentioned in Table 4. The approach is validated using k-fold cross validation with $k = 10$ folds is fixed for each techniques. The dataset is divided into 10 equal parts and in each iteration 9 sets are used to train the algorithm and 1 set is used to test learned model.

After observing the results mentioned in Table 4, we can easily conclude that the decision tree based method has performed well for the proposed technique. The F1-score score obtained using Random Forest is 0.918. Random forest is a powerful classifier based on the decision tree algorithm and its voting method selects the best performing classifier based on bootstrapping applied to the relevant dataset. The KNN algorithm performs a little worse than decision tree with an F1-score score as 0.86. However, Naive Bayes performance is quite low as compared to other classifiers with the F1-score score of 0.83.

Table 3
List of original features [15] .

List of original features
BPS-24, BPD-24, BPS-Day24, BPD-Day24, BPS-Night24, BPD-Night24, BPS-load-Day, BPD-load-Day, BPS-load-Night, BPD-load-Night, Max-Sys, Min-Sys, Max-Dia, Min-Dia, BPS-CV-all, BPD-CV-all, BPS-CV-Day, BPD-CV-Day, BPS-CV-Night, BPD-CV-Night, BPS-wakeUp, BPD-wakeUp, low-BPS-Night, low-BPD-Night, Age, Sex, Height, Weight

Table 4
Performance evaluation metric for multi-label classification.

Algorithm	Precision	Recall	F1	Hamming Loss	Jaccard Score	Parameters
Naïve Bayes	0.82	0.856	0.835	0.183	0.72	-
Random Forest	0.903	0.933	0.918	0.127	0.75	estimators=20, criterion= gini
KNN	0.879	0.855	0.86	0.157	0.74	k=5, metric= Minkowski, p=1.5

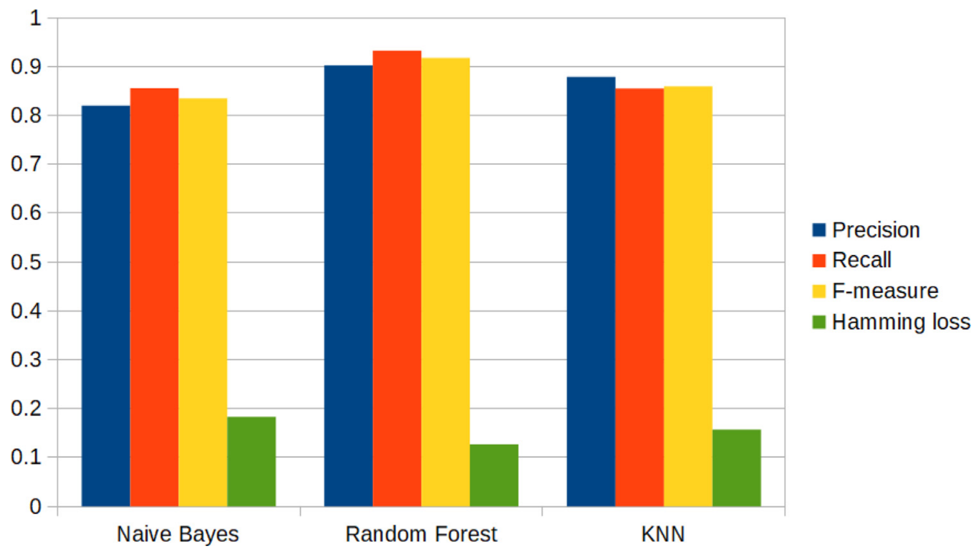

Fig. 2. Graphical representation of the obtained results.

4. Discussion

Knowledge-driven feature engineering has been used in several health-related studies in the literature. J. Nahar et al. have presented a comparison between an automated feature selection method and knowledge-based feature selection method and have shown that knowledge based feature selection method improves the performance of the prediction model in terms of accuracy (up to 97% in some cases) [21]. Knowledge-driven feature engineering is used in various medical domains including cancer diagnosis, radiography, ultrasound, MRI, clinical text reports, and sensor-based health monitoring in general. For example, H. Feng et al. have used domain knowledge to guide the feature learning process for deep learning models in breast cancer diagnosis [22]. Often limited amount of labelled data is available in the medical domain because a smaller number of participants are willing to share their personal data due to privacy reasons. In such cases, prior knowledge about the feature-label distribution is helpful to improve the accuracy of the model even having a small amount of data available for training purpose. S. Boluki et al. proposed a method which uses biological prior knowledge of feature-label distribution to develop an optimized Bayesian classification model for gene regulatory data [23]. A. Wilcox et al. have studied the effects of expert knowledge on inductive learning for medical text report classification and found that acquired knowledge significantly contributes to the performance of clinical classification task [24], where the predictive models comparatively performed worst when domain knowledge is not considered ($p < .001$). In another case study, authors studied the effect of including knowledge into the feature engineering process, prior to applying machine learning models on the data acquired from severe asthma patients [9]. The study concludes that incorporating knowledge has reduced the complexity of the computational model and also improved the performance of the predictive model, where the change in area under the curve $\triangle AUC$ was ≤ 0.03 obtained from all the considered modelling approaches.

The second objective of our proposed model is to use knowledge-driven features to solve a multi-label classification [25] problem in which multiple indicators are associated with the extracted knowledge-based features. Multi-label classification is useful in the medical domain as often the available clinical data reflects multiple symptoms required to diagnose a disorder. In the medical domain, multi-label classification has been widely used to classify text and image data linked to multiple indicators. For example, J. Du et al. proposed ML-Net, a deep learning based framework for multi-label classification of biomedical text [26]. H. Chougrad et al. have used multi-label image classification to diagnose early stages of breast cancer [27]. However, limited work has been done in cardiovascular and blood pressure disorder detection using multi-label classification. Multi-label classification for electrocardiography (ECG) data was presented by Z. Sun et al. in which an ensemble classifier is applied on the multi-label ECG data [28]. In another work K. Doubi et al. have applied multi-label classification on ABPM dataset to detect multiple symptoms [15] with an average $F1 - score = 0.92$. They presented their results for many multi-label classification techniques such as binary relevance, label powerset, and RAndom k-labELsets (RAkEL), while using decision tree as the base algorithm for classification. This research work provides the basics for our proposed model. Their work is based on data-driven feature extraction, whereas our proposed model relies on knowledge-driven features for the detection of multiple hypertension symptoms. The obtained results show that, with knowledge-driven features, we can achieve an optimum detection performance by only using 8 features, in comparison to 40 features used in the original dataset. Douibi et al. [15] et al. reported an accuracy per label for circadian rhythm, pulse pressure, and morning surge of 0.985, 0.856, and 0.837 using RAKEL, as well as general hamming loss of 0.064 and F1 (micro averaged) of 0.953. Our objective is to show that, by including a knowledge-driven feature engineering step in the learning process, the feature-set could be significantly reduced (5-times compared to the original feature-set, in the studied case [15]) without a major impact on the overall performance. The *F1-score* scores show that knowledge-driven features has given satisfactory performance for multi-label classification. Our approach transforms the numeric hypertension measurements into an understandable format, which not only plays a role in "correct" learning of the model but also provides clinicians with an abstract representation of the features for the better understanding and fine-tuning of the model. Knowledge-driven feature extraction could also reduce the complexity of learning as evident from the results that we have achieved almost similar performance while using 8 knowledge-driven features as compared to 40 data-driven features. Often feature reduction is required to implement the technique on resource constrained devices such as a micro-controller and a raspberry-pi for a wearable edge device. Hence, with *knowledge enriched features*, we can achieve a good performance in terms of feature interpretability (or abstract representation of data for better understanding for clinicians), machine learning complexity, feature reduction, reduced learning time, and most important embedding knowledge to obtain a clinical correct/fine-tuned model to detect multiple symptoms. We can further enhance the detection performance of the model by adding relevant knowledge, however, it depends on the availability of data as well as domain experts' knowledge. While the application of knowledge-driven feature engineering for medical applications has been considered in some case studies in literature (as mentioned in Section 4), this is the first time that this approach, as well as multi-label classification, is applied together to a blood pressure-related study.

This study has some limitations related to acquiring domain knowledge, quality of available data, increase in model complexity as the number of descriptive features increases, and time consumption in manual rule formulation. Our proposed approach is based on manual knowledge acquisition which gives us precision, however, the overall process becomes tedious as the number of descriptive features increases. An extension of this work is possible by applying automatic rule extraction methods for text mining on reliable medical knowledge sources. Text mining can automatically discover a range of rules to produce new knowledge-driven features. However, quality of these rules must be examined by a medical expert to ensure robust knowledge-driven features for the training purpose.

5. Conclusion

In conclusion, we have shown that the proposed approach based on knowledge-driven feature engineering and multi-label classification could correctly classify multiple symptoms related to blood pressure measurements. This proposed approach may be useful in tackling several problems in medical prediction models, such as white-box transparency, feature reduction, fine-tuning of descriptive features by including experts' knowledge in the detection process, and allowing the model to detect multiple disorders. We have achieved satisfactory classification results for three conditions: circadian rhythm, pulse pressure, and morning surge using Random forest, KNN, and Naive Bayes techniques. The detection of each of these may translate into improved clinical assessment of patients with hypertension, in the detection of the presence of hypertension, in the assessment of the adequacy of the therapeutic response to lifestyle and medications, and in the prevention of the complications associated with hypertension.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript:

Zaffar Haider Janjua, Tyndall National Institute, Cork, Ireland. Davide Kerins, University College, Cork, Ireland. Brendan Olynn, Tyndall National Institute, Cork, Ireland. Salvatore Tedesco, Tyndall National Institute, Cork, Ireland.

Acknowledgements

This publication has emanated from research supported by a research grant from the Disruptive Technologies Innovation Fund (DTIF) project HOLISTICS funded by Enterprise Ireland (EI). Aspects of this publication have emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 12/RC/2289-P2 INSIGHT, 13/RC/2077-CONNECT, and 16/RC/3918-CONFIRM, which are co-funded under the European Regional Development Fund.

References

[1] WHO, Cardiovascular diseases (CVDs) fact sheet, 11 June 2021, Accessed on 10/11/2021, https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
[2] M.B. Liu, Cardiovascular diseases, Chin Med J 127 (2014) 6–7.
[3] T. Unger, C. Borghi, F. Charchar, N.A. Khan, N.R. Poulter, D. Prabhakaran, A. Ramirez, M. Schlaich, G.S. Stergiou, M. Tomaszewski, R.D. Wainford, B. Williams, A.E. Schutte, 2020 international society of hypertension global hypertension practice guidelines, Hypertension (2020) 1334–1357, doi:10.1161/HYPERTENSIONAHA.120.15026.
[4] A. Dadlani, K. Madan, J.P.S. Sawhney, Ambulatory blood pressure monitoring in clinical practice, Indian Heart J 71 (1) (2019) 91–97, doi:10.1016/j.ihj.2018.11.015.
[5] E.A. Andreadis, C.V. Geladari, E.T. Angelopoulos, F.S. Savva, A.I. Georgantoni, V. Papademetriou, Attended and unattended automated office blood pressure measurements have better agreement with ambulatory monitoring than conventional office readings, J Am Heart Assoc 7 (8) (2018), doi:10.1161/JAHA.118.008994.
[6] D. Shimbo, M. Abdalla, L. Falzon, R.R. Townsend, P. Muntner, Studies comparing ambulatory blood pressure and home blood pressure on cardiovascular disease and mortality outcomes: a systematic review, Journal of the American Society of Hypertension 10 (3) (2016) 224–234.e17, doi:10.1016/j.jash.2015.12.013.
[7] M.J. Zawadzki, A.K. Small, W. Gerin, Ambulatory blood pressure variability, Blood Press Monit 22 (2) (2017) 53–58, doi:10.1097/MBP.0000000000000230.
[8] D.S. Char, N.H. Shah, D. Magnus, Implementing machine learning in health care – addressing ethical challenges, N top N. Engl. J. Med. 378 (11) (2018) 981–983, doi:10.1056/NEJMp1714229.
[9] K.D. Roe, V. Jawa, X. Zhang, C.G. Chute, J.A. Epstein, J. Matelsky, I. Shpitser, C.O. Taylor, et al., Feature engineering with clinical expert knowledge: a case study assessment of machine learning model complexity and performance, PLoS ONE 15 (4) (2020), doi:10.1371/journal.pone.0231300.
[10] J. Read, A. Puurula, A. Bifet, Multi-label classification with meta-labels, in: 2014 IEEE International Conference on Data Mining, 2014, pp. 941–946.
[11] G. Tsoumakas, I. Katakis, I. Vlahavas, Random k-labelsets for multilabel classification, IEEE Trans Knowl Data Eng 23 (7) (2011) 1079–1089, doi:10.1109/TKDE.2010.164.
[12] G. Mancia, G. De Backer, A. Dominiczak, R. Cifkova, R. Fagard, G. Germano, G. Grassi, A.M. Heagerty, S.E. Kjeldsen, S. Laurent, K. Narkiewicz, L. Ruilope, A. Rynkiewicz, R.E. Schmieder, H.A.J.S. Boudier, A. Zanchetti, 2007 guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the european society of hypertension (ESH) and of the european society of cardiology (ESC), J. Hypertens. 25 (6) (2007) 1105–1187, doi:10.1097/HJH.0b013e3281fc975a.
[13] P.K. Zachariah, S.G. Sheps, D.M. Ilstrup, C.R. Long, K.R. Bailey, C.M. Wiltgen, C.A. Carlson, Blood pressure loada better determinant of hypertension, Mayo Clin. Proc. 63 (11) (1988) 1085–1091, doi:10.1016/S0025-6196(12)65503-7.
[14] C.D. Manning, P. Raghavan, H. Schutze, Introduction to information retrieval, 2008. Cambridge University Press Cambridge, England
[15] K. Douibi, N. Settouti, M.A. Chikh, J. Read, M.M. Benabid, An analysis of ambulatory blood pressure monitoring using multi-label classification, Australasian Physical and Engineering Sciences in Medicine 42 (1) (2019) 65–81, doi:10.1007/s13246-018-0713-0.
[16] Y.F. Guo, P.K. Stein, Circadian rhythm in the cardiovascular system: chronocardiology, Am. Heart J. 145 (5) (2003) 779–786, doi:10.1016/S0002-8703(02)94797-6.
[17] T.D. Homan, S. Bordes, E. Cichowski, Physiology, Pulse Pressure - StatPearls - NCBI Bookshelf, 2018, Accessed on 19/08/2020. https://www.ncbi.nlm.nih.gov/books/NBK482408/

[18] K. Kario, Morning surge in blood pressure and cardiovascular risk: evidence and perspectives, Hypertension (Dallas, Tex. : 1979) 56 (5) (2010) 765–773, doi:10.1161/HYPERTENSIONAHA.110.157149.
[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.
[20] P. Szymański, T. Kajdanowicz, A scikit-based Python environment for performing multi-label classification, ArXiv e-prints (2017).
[21] J. Nahar, T. Imam, K.S. Tickle, Y.P.P. Chen, Computational intelligence for heart disease diagnosis: a medical knowledge driven approach, Expert Syst Appl 40 (1) (2013) 96–104, doi:10.1016/j.eswa.2012.07.032.
[22] H. Feng, J. Cao, H. Wang, Y. Xie, D. Yang, J. Feng, B. Chen, A knowledge-driven feature learning and integration method for breast cancer diagnosis on multi-sequence MRI, Magn Reson Imaging 69 (2020) 40–48, doi:10.1016/j.mri.2020.03.001.
[23] S. Boluki, M.S. Esfahani, X. Qian, E.R. Dougherty, Incorporating biological prior knowledge for bayesian learning via maximal knowledge-driven information priors, BMC Bioinformatics 18 (14) (2017) 61–80, doi:10.1186/s12859-017-1893-4.
[24] A.B. Wilcox, G. Hripcsak, The role of domain knowledge in automating medical text report classification, Journal of the American Medical Informatics Association 10 (4) (2003) 330–338, doi:10.1197/jamia.M1157.
[25] F. Herrera, F. Charte, A.J. Rivera, M.J. del Jesus, Multilabel Classification : Problem Analysis, Metrics and Techniques, 2016. Springer International Publishing
[26] J. Du, Q. Chen, Y. Peng, Y. Xiang, C. Tao, Z. Lu, et al., ML-Net: Multi-label classification of biomedical texts with deep neural networks, Journal of the American Medical Informatics Association 26 (11) (2019) 1279–1285, doi:10.1093/jamia/ocz085.
[27] H. Chougrad, H. Zouaki, O. Alheyane, Multi-label transfer learning for the early diagnosis of breast cancer, Neurocomputing 392 (2020) 168–180, doi:10.1016/j.neucom.2019.01.112.
[28] Z. Sun, C. Wang, Y. Zhao, C. Yan, Multi-Label ECG signal classification based on ensemble classifier, IEEE Access 8 (2020) 117986–117996, doi:10.1109/ACCESS.2020.3004908.