Expert Systems With Applications 239 (2024) 122388

Contents lists available at ScienceDirect

Expert Systems With Applications

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

ExHyptNet: An explainable diagnosis of hypertension using EfficientNet with PPG signals

El-Sayed A. El-Dahshan [a,b,*], Mahmoud M. Bassiouni [b], Smith K. Khare [c], Ru-San Tan [d,e], U. Rajendra Acharya [f,g]

[a] Department of Physics, Faculty of Science, Ain Shams University, Postal Code: 11566 Cairo, Egypt
[b] Egyptian E-Learning University (EELU), 33 El-messah Street, Eldoki, Postal Code: 11261 El-Giza, Egypt
[c] Applied AI and Data Science Unit, Mærsk Mc-Kinney Møller Institute, Faculty of Engineering, University of Southern Denmark, Denmark
[d] Department of Cardiology, National Heart Centre Singapore, Singapore
[e] Duke-NUS Medical School, Singapore
[f] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia
[g] Centre for Health Research, University of Southern Queensland, Australia

ARTICLE INFO

Keywords:
PPG signals
Hypertension
Recurrence plots
EfficientNetB3
Explainable AI

ABSTRACT

Background: Hypertension is a crucial health indicator because it provides subtle details about a patient's cardiac health. Photoplethysmography (PPG) signals are a critical biological marker used for the early detection and diagnosis of hypertension.
Objective: The existing hypertension detection models cannot explain the model's prediction, making it unreliable for clinicians. The proposed study aims to develop an explainable and effective hypertension detection (ExHyptNet) model using PPG signals.
Methods: The proposed ExHyptNet model is an ensemble of multi-level feature analyses used to detect and explain hypertension predictions. In the feature extraction stage, recurrence plots and EfficientNetB3 architecture are employed to extract deep features from the PPG signals. Then, features are explained using a Gradient-weighted Class Activation Mapping (Grad-CAM) explainer in the explainable stage. In the last stage, XG-Boost and extremely randomized trees (ERT) classifiers are used to make the qualitative and quantitative analysis for evaluating the performance of the proposed ExHyptNet model.
Results: The performance of the ExHyptNet model is evaluated on two public PPG datasets: PPG-BP and MIMIC-II, using holdout, stratified 10-fold cross-validation, and leave-one-out subject validation techniques. The developed model yielded a 100% detection rate for the classification of normal and multi-stage hypertension classes using three validation techniques. The proposed work also demonstrates a detailed ablation study using hyper-parameters, pre-trained models, and the detection of several PPG categories.
Conclusion: The developed ExHyptNet model performed better than the existing automated hypertension detection systems. Our proposed model is practically realizable to clinicians in real-time hypertension detection as it is validated on two public PPG datasets using different validation techniques.

1. Introduction

Blood pressure (BP) is an important cause of death and morbidity globally. Hypertension termed a silent killer is a key risk factor along with other (Cardiovascular diseases) CVDs like atrial fibrillation (Bassiouni et al., 2018) and other risk factors like high cholesterol (Loh et al., 2022; W.H. Organization, 2013; Al-Zaben et al., 2018). Untreated or delayed hypertension may arouse risk factors linked to heart, brain (El-Dahshan & Bassiouni, 2018), and kidney diseases (Hall, 2003; Drozdz & Kawecka-Jaszcz, 2014) Timely diagnosis and medication of hypertension can avoid the risk factors of many diseases, saving many lives(Zhu et al., 2019). Measurement of BP in millimeters of mercury (mmHg) involves three vital parameters namely: mean arterial pressure (MAP), systolic blood pressure (SBP), and diastolic blood pressure (DBP),

* Corresponding author.
E-mail addresses: seldahshan@sci.asu.edu.eg (E.-S.A. El-Dahshan), mbassiouni@eelu.edu.eg (M.M. Bassiouni), smkh@mmmi.sdu.dk (S.K. Khare), Rajendra.Acharya@usq.edu.au (U. Rajendra Acharya).

https://doi.org/10.1016/j.eswa.2023.122388
Received 20 April 2023; Received in revised form 10 October 2023; Accepted 27 October 2023
Available online 2 November 2023

respectively (Azadifar et al., 2022). A healthy individual's DBP and SBP of less than 120 mmHg and 90 mmHg are considered normal. The SBP and DBP in 120–139 and 80–89 mmHg are categorized as pre-hypertension. SBP between 140 and 159 and DBP between 90 and 99 mmHg is termed stage 1 hypertension, while SBP higher than 160 and DBP higher than 100 mmHg is stage 2 hypertension, respectively (Maqsood et al., 2021).

Photoplethysmography (PPG) is a promising technique for hypertension detection (Almarshad et al., 2022). PPG devices access the changes in blood volume in the microcirculation of an accessible body part e.g. wrist or pulp of fingertip, using a light-emitting diode and photodetector. PPG works by shining light onto the skin and measuring the changes in light absorption caused by the pumping of blood through the underlying blood vessels. The signal is used to study many physiological parameters such as heart rate, blood pressure, and blood flow. The advantage of PPG acquisition is that the measurement of PPG signals offers cost-effective, fast, non-invasive, and easy solutions. Wearable smartwatches and smartphones have been developed that can measure BP using PPG (Azadifar et al., 2022). However, these devices are mostly consumer-grade and are not clinically accepted due to their limited accuracy for BP measurement compared with standard cuff-based and invasive methods (Scheer et al., 2002; Tjahjadi & Ramli, 2020) Nevertheless, PPG remains promising for classifying the different stages of BP (Tjahjadi & Ramli, 2020). Its widespread availability makes it a cost-effective technique for broad population screening of hypertension. Its ability for continuous signal tracking offers efficient remote monitoring in diverse clinical Applications, e.g. cardiac rehabilitation.

2. Related work

Many PPG-based automated hypertension stage classification systems have been developed (Table 1). Most models are developed on a single PPG dataset, either on the public or private dataset.

i. The existing HYPT detection systems use random validation strategies limited to either holdout or k-fold Cross-validation (CV) techniques.
ii. The models have limited performance due to limited qualitative and quantitative analysis.
iii. Clinicians resist the existing automated systems due to a lack of explainability.

Due to the limitations mentioned above of the existing hypertension detection systems, the following research questions (RQs) are still open:

RQ1. Generalization: The model must be robust to deploy the automated systems in the clinical setting.

Table 1
Summary of works done on PPG-based machine learning algorithms for hypertension classification.

Study	Dataset	Subjects	Feature	Validation	Classifier	Acc (%)	Sen (%)	Spe (%)
Hettiarachchi and Chitraranjan (2019)	Public clinical	51 HC, 39 PH, 28 HT, 9 DM, 16 DPH, 7 DHT	Skewness signal quality index	10-fold CV	Decision tree	83	–	–
Mousavi et al., 2019	MIMIC-II	441 subjects; 1323 records	Whole-based feature	10-fold CV	SVM	–	–	–
Yao and Liu, 2021	Public clinical	51 HC, 39 PH, 28 HT, 9 DM, 16 DPH, 7 DHT	Complete ensemble empirical mode decomposition with adaptive noise	10-fold CV	SVM	89	81	85
Cano et al., 2021	MIMIC-III	50 subjects; 635 records	CWT	Holdout	CNN	–	93.1	76.0
Nour & Polat, 2020	PPG-BP	219 subjects; 657 records	Physiological features	5-fold CV	Random forest	99.5	–	–
Khan et al. (2021)	Private	121 subjects; 700 HT records, 709HC records	Empirical mode decomposition	20-fold CV	K-nearest neighbors	99.7	99.2	99.4
Evdochimet al., 2022	PPG-BP	219 subjects	Signal morphology	–	Decision tree	69.9	66.7	71.1
	MIMIC-III	140 subjects			SVM	72.9	71	67.5
Gupta et al. (2022)	PPG-BP	219 subjects; 657 records	Filtering	–	CNN-LSTM (Long-short	67.8	68.4	66.6
Khan et al. (2022)	PPG-BP	219 subjects; 657 records	Variational mode decomposition	10-fold CV	Gradient boosting classifier	99.3	98.7	100
Ranjan et al., 2022	PPG-BP	219 subjects; 657 records	Filtering	–	CNN	74.5	73.4	75.7
Sadadet al., 2022	PPG-BP	219 subjects; 657 records	Heart rate	k-fold CV	Decision tree	99.5	–	–
	PPG-DaLiA	2 subjects	–	Holdout	CNN-LSTM	97.6	–	–
Tanc and Ozturk (2022)	MIMIC-II	150 records	Synchrosqueezing transform	Holdout	GoogLeNet	–	95.8	96.0
Gupta et al. (2022)	MIMIC-I MIMIC-II MIMIC-III	39 subjects 120,000 records 510 subjects	Tunable q-factor wavelet transform	–	Random forest	–	–	–
Wu et al., 2021	MIMIC-III	30,000 subjects; 67,830 records	CWT	Holdout	CNN	90	–	–
Liang et al., 2018	MIMIC	121 subjects	CWT	Holdout	CNN	92.6	–	–
Yen et al., 2021	PPG-BP	219 subjects; 657 records	–	Holdout	ResNetCNN	73	–	–
Nafisi and Shahabi (2018)	Private	10 subjects	Statistical features	10-fold CV	AdaBoost	96.6	94.9	97.4
Mejía-Mejía et al. (2021)	MIMIC-II	32,536 subjects	Nonlinear features	10-fold CV	SVM	70	50	75
Lee et al. (2021)	VitalDB	3301 subjects	Physiological parameters	Holdout	CNN		80.7	80.7
Ardinyet al., 2015	WESAD	17 subjects	–	Holdout	CNN	96.7		
Sannino et al., 2020	MIMIC-II	526,906 instances	–	Holdout	Random forest	99.4	–	–
Lan et al., 2018	Private	43 Subjects	Physiological parameters based on SDNN of HRV features	Holdout	MIL	85.4	–	–

Acc-Accuracy; Rec-Recall; Spe-Specificity; CV-cross-validation; CNN-convolutional neural network; CWT-continuous wavelet transforms; DM-diabetes; DHT- diabetic hypertension; DPH-diabetic prehypertension; HC-healthy control; HT-hypertension; LOOCV-leave-one-out cross-validation; LSTM-long short-term memory; PH-prehypertension; SVM-support vector machine; VMD- variational mode decomposition; SST-Synchrosqueezing transform; Standard deviation of NN, MIL- Multiple-Instance Learning.

RQ2. Explainability: Due to the lack of explainability, clinicians find it difficult to trust the predictions given by the existing ML and DL models.

In this paper, we have tried to narrow the above-mentioned research gaps by developing a robust, effective, accurate, explainable, and versatile model. Our developed automated and explainable hypertension detection (ExHyptNet) model developed using PPG signals has the following contributions:

i. The ExHyptNet combines recurrence plots (RP), EfficientNetB3, and an extremely randomized tree (ERT) classifier tested on two public PPG datasets.
ii. The model is developed using three validation techniques: holdout, stratified 10-fold CV, and leave-one-out subject validation (LOOSV) techniques.
iii. The model explains the regions of the RP responsible for the prediction of hypertension.
iv. The accuracy of the model is tested by qualitative (performance measures) and quantitative (Taylor, Box, and Spider plots) analyses.

The remainder of the paper is organized as follows: Section 3 presents the proposed methodology, results are presented in Section 4, Discussion on the results is given in Section 5, and finally, the paper concludes in Section 6.

3. Methodology

A novel methodology is proposed to diagnose hypertension based on PPG signals based on a set of phases. The first phase is based on data acquisition, and in this phase, the PPG signals are obtained from two main datasets: PPG-BP (Photoplethysmography- Blood pressure dataset) and the MIMIC database. The second phase relies on converting the 1D PPG signals into 2D images based on the RP (Marwan et al., 2007). The third phase is the feature extraction phase, and in this phase, the PPG recurrence images are fed to a pre-trained model based on EfficientNetB3 for feature extraction. The fourth phase relies on explaining the important features and components in the PPG RP images using explainable AI techniques. In the fourth phase, the Grad-CAM technique (Selvaraju et al., 2017) is used. Finally, the fifth phase is the classification phase which is done using Extreme Gradient Boosting (XGBoost) (Chen et al., 2015) and ERT (Geurts et al., 2006) classifiers. It is important to mention that several qualitative and quantitative measurements are provided on the results obtained from the entire methodology. Fig. 1 shows the overall methodology used for diagnosing hypertension based on PPG signals.

3.1. PPG datasets

We downloaded PPG signals from 2 public databases: PPG-BP (Liang et al., 2018) and MIMIC (Lee et al., 2011). Table 2 shows a full description of the two datasets. The first dataset consists of four main classes: normal, pre-hypertension, stage1 hypertension, and stage2 hypertension. The number of subjects in this dataset is 219 adults, and the subjects are aged 21–86 years. Also, for each subject 8 main features are defined, such as sex, age, height, weight, systolic blood pressure, diastolic blood pressure, heart rate, and body mass index (BMI). Each subject consists of 3 signals forming a total of 657 signals. The second dataset is the MIMIC database, an available online dataset. It consists of thousands of patients. Several physiological signals exist in this dataset: atrial blood pressure (ABP), electrocardiogram (ECG), and PPG. Some signals hold noisy, unclean, missing peaks, and pulse bisferiens. Therefore, these signals are removed completely and not selected. 121 subjects are collected, each with a single PPG signal. The subjects are divided into three categories: normal, prehypertension, and hypertension. In the PPG-BP and MIMIC-II datasets, the Systolic and Diastolic Blood Pressures are calculated for each PPG category. Table 3 shows the average of the systolic and diastolic blood pressure values for each category in the two datasets. In the PPG dataset, each signal is assigned to a specific blood pressure value and to a group category (normal, hypertension, and prehypertension). For instance, subject number two is a female and all its signals are assigned to level 2 hypertension.

3.2. Preprocessing using recurrence plots

The PPG signals from both datasets are de-noised using a 9th-order Butterworth low pass filter with a cut-off frequency of 500 Hz and the signals are smoothed using a Savitzky–Golay filter with a smoothing value of 5. Each record in the PPG-BP and MIMIC-II datasets consists of 2100 and 8200 samples, respectively. One of the powerful tools used for

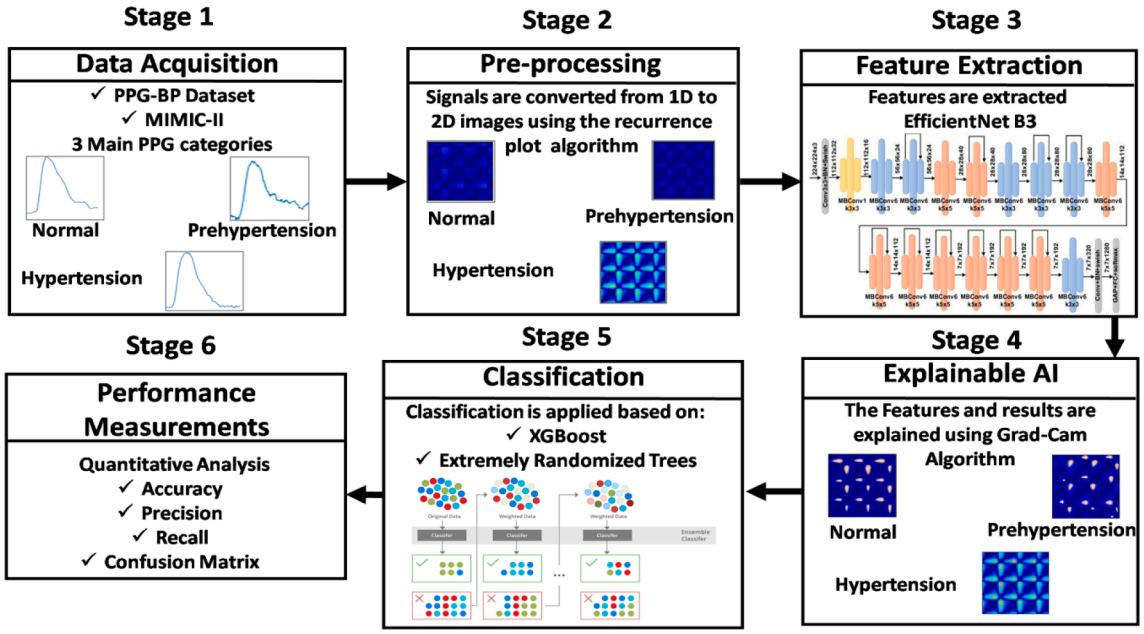

Fig. 1. Proposed ExHyptNet for automated detection of hypertension using PPG signals.

Table 2
Description of PPG-BP and MIMIC-II datasets.

Dataset Name	No of Subjects	No PPG Signal	Segment Sample	Age	Sampling Rate	Time duration
PPG-BP	219	657 Each Signal consists of 3 PPG segments	Each segment is 2100 sampling point	21–86 48 % of them are male	1 kHz, with 12 bits ADC conversion precision	Each PPG segment is 2.1 Seconds
MIMIC-II	26,870 121 subjects have PPG signals	67,830 121 Signals Has hypertension	Each signal is 8200 sampling point	65–80	125 Hz with 12 bits ADC conversion precision	Each PPG signal is 120 s in length

Table 3
Average systolic and diastolic blood pressure values for each category in the two PPG datasets.

Dataset Name	Normal		Prehypertension		Stage 1 Hypertension		Stage 2 Hypertension	
	SBP	DBP	SBP	DBP	SBP	DBP	SBP	DBP
PPG-BP	128	72	129	70	**145**	**86**	**162**	**98**
MIMIC-II	129	75	129	79	**142**	**88**	**165**	**96**

analyzing the data is the RPs. RPs do not visualize the structure of the time series, but they estimate the invariants resulting in them (Lee et al., 2019). Also, these plots can analyze the non-stationary data and quantify any hidden structures in the data. These plots are very beneficial in linear dynamic systems analysis, even when combined with non-linear features. An RP can be used to analyze physiological signals that can be electrical, chemical, mechanical, hormonal, and neural (Roh & Shin, 2021). The main aim of the RP is to visualize the recurrence state in phase space. This procedure permits the representation of a 1D-dimensional PPG signal in a 2D dimension based on its recurrence. The recurrence between time i and time j is defined using a two-dimensional array (Jan et al., 2019). Let us consider the PPG signal in the following sequence:

$$S = \{S_1, S_2, S_3, \cdots..S_n\} \tag{1}$$

Then, based on the sequence there exists a d-dimensional vectors a_k that can be constructed using a time-delay procedure embedding. The vectors of a_k are defined using the following equation:

$$a_k = \left\{S_k, S_{k+\tau}, S_{k+2\tau}, \cdots.., S_{k+(d-1)\tau}\right\} \tag{2}$$

where τ is the delay parameter and d is defined as the embedding dimension. The parameter d is optimized using the false nearest neighbour's (FNN) method. Each dimension has a time unit with values of 1 and 0. The recurrence between time i and time j is represented using the following equation:

$$R_{i,j} = \Theta\left(\varepsilon_i - \left|\left|\overline{S}_{i+k} - S_{j+k}\right|\right|\right), \overline{S}_i \in \mathbb{R}^m, i,j = 1, \cdots\cdots, N \tag{3}$$

The number of states S_i are represented by N, ε_i and $\|.\|$ represent the threshold distance and the norm, respectively. m refers to m-dimensional phase space trajectory through a two-dimensional representation of its recurrences. ε is the threshold distance parameter and its value must be greater than the separation performed on each pair of coordinates in the series. Finally, $\Theta(.)$ is known as the unit step function. In this study, each PPG signal was converted to an RP to produce a two-dimensional image that can be input into CNN architecture. Since each PPG signal can have a different number of samples the produced RPs are interpolated using a bicubic interpolation (Barroso-García et al., 2020). The aim is to obtain a fixed recurrence image dimension for each PPG signal of size 124 x 124. Fig. 2(a, b, and c) shows the RP obtained for normal, prehypertension, and hypertension, respectively. It is also important to mention that in the PPG-BP each PPG signal is 2.1 s long, while for the MIMIC-II dataset, each PPG signal is 120-second-long. Then, each PPG signal in each dataset is entirely converted from the 1D signal to obtain one 2D RP image. Finally, it is important to mention that the characteristic topology of the presented RPs is periodic and contains periodic recurrent structures (vertical and horizontal lines) representing the main variances.

3.3. Features extraction using pre-trained EfficientNetB3

(Tan & Le, 2019) conducted a study to establish the relationship between the depth and width of CNN models. The outcome was the creation of EfficientNet CNN models, which can achieve high classification accuracy with fewer parameters. These models range from EfficientNet-B0 to EfficientNet-B7 and outperform pre-trained models in terms of parameters and top-1 accuracy on the ImageNet dataset. EfficientNet scales each dimension uniformly, but balancing the dimensions can improve overall performance. The EfficientNetB3 CNN model is used as it strikes a balance between accuracy and computational resources. The Mobile Inverted Bottleneck Convolution (MBConv) block is

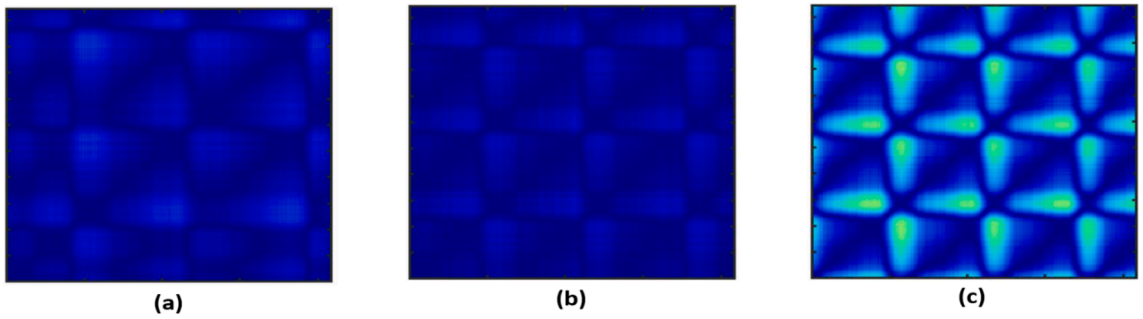

Fig. 2. Typical RPs: (a) normal, (b) prehypertension, and (c) hypertension.

the backbone of the EfficientNet family, which is derived from the MobileNet pre-trained model. The MBConvBlock comprises five aspects: depth-wise separable convolution layers, inverted residual connections, linear bottlenecks, squeeze-and-excitation, and swish activation functions. The MBConvBlock is defined by two types: MBConv6 and MBConv1, with different input feature map sizes and filter sizes. The EfficientNetB3 model is trained on two PPG datasets, with the data split into 60 % training, 20 % validation, and 20 % for testing.

3.4. Explainable AI (XAI) using gradient class activation mapping (Grad-CAM)

Explainable artificial intelligence (XAI) (Kashevnik and Ali, 2022) refers to the processes and methods that enable humans to understand and trust the results and outputs produced by machine learning algorithms (Ganesh et al., 2021; Angelov et al., 2021). XAI aims to provide insight into an AI model's potential biases, expected impact, accuracy, fairness, transparency, and outcomes in decision-making. One type of XAI is Grad-CAM, proposed by (Selvaraju et al., 2017; Selvaraju et al., 2016). Grad-CAM can produce high-resolution and class-discriminative Guided Grad-CAM visuals to explain how deep learning models generate their most discrete features. This is achieved by combining current pixel-space gradient visualizations, revealing significant parts of an image that correspond to any decision of interest. Grad-CAM helps understand and identify the attention of a CNN model for classification. It computes weights generated from each feature map concerning the scores obtained from the classification process to produce an attention or heat map (Chen et al., 2015). Grad-CAM is an advanced version of class activation mapping (CAM), with the attention map of CAM serving as the starting point for Grad-CAM. It's worth noting that the CNN model needs to be trained before applying Grad-CAM (Loh et al., 2022). The primary advantage of Grad-CAM is that the model's main structure does not affect the process, as it relies on feature maps and classification results.

3.5. Classification

To make accurate decisions based on a proposed methodology system, classification is necessary. There are various classifiers available for this purpose, but this study focuses on two main ones: XGBoost and ERT learning. These classifiers will be used to classify the proposed methodology.

3.5.1. Extreme Gradient boosting (XGBoost) classifier

XGBoost is a powerful and scalable machine learning system for tree boosting that has been widely utilized in various fields to achieve state-of-the-art results in many data challenges (Friedman, 2001). Chen and Guestrin introduced XGBoost in 2016 to enhance classification performance. In the 2015 Kaggle machine learning competition, XGBoost was the most commonly used method, and it was employed in 17 of the 29 winning solutions. The features of the EfficientNetB3 were trained using XGBoost because of its excellent performance in supervised machine learning. The XGBoost classifier is an ensemble of trees that combines multiple classification and regression trees.

3.5.2. Extremely randomized trees (ERT) classifier

ERT is a classifier that is similar to random decision forests, but it introduces randomness during the training process (Saeed et al., 2021). ERT uses an ensemble of highly random trees that are induced on the learning set of sub-windows to perform the classification. At each node in the tree creation process, ERT follows three steps. Firstly, it selects K distinct features randomly from the entire feature vector. Secondly, it computes one random split for each feature. Finally, it chooses the split that maximizes a score measure based on a normalized version of the Shannon information gain (Sharaff & Gupta, 2019). K is a parameter that determines the degree of randomization used in tree induction, and a high value of K filters out irrelevant variables, making the tree classification process more accurate. However, it also introduces more correlation between trees in the forest. Two other parameters, n estimators and *nmin*, affect the performance of the extra-trees. The number of trees in the ensemble, L, affects the intensity of variance reduction, while *nmin* governs the strength of averaging output noise. These parameters can be adjusted automatically or manually to suit the unique characteristics of the input features.

4. Results

The experimental setup, hyperparameters used, classifier selection, the model used, and classification results are presented below.

4.1. Experimental setup

The experiments were implemented in the MATLAB R2022a programming environment on a laptop computer with an Intel Core i7-8565 central processing unit, 12 Gigabyte random access memory, and an NVIDIA GeForce MX 130 M 2 GB graphical processing unit. In total, 6 experiments were performed for multi-class (2) and binary (4) classifications, evenly divided between the PPG-BP (3) and MIMIC-II (3) datasets. The training: validation: and test data split in each experiment was 60 %: 20 %: 20 %, respectively. Standard metrics were used to evaluate performance: true (TP) and false positive (FP) rates, precision (Pre), recall (Rec), Matthew's correlation coefficient (MMC), receiver operating characteristic analyzed area under the curve (ROC), performance curve (PRC), (García et al., 2009) kappa statistic (κ), and accuracy (Acc %). Confusion matrices were constructed. Finally, t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten 2008) was applied to the multi-class classification experiments to depict how separable the PPG features were by the various classifiers.

4.2. Hyperparameters and classifiers settings

Hyperparameter settings of the EfficientNetB3 deep learning model used in the study were set after preliminary experiments using the adaptive moment estimation optimizer (Adam) and are shown in Table 4. The parameters of the classifiers are defined in Table 5.

Algorithm 1 shows the main stages and steps of the ExHyptNet in terms of RP calculation, EfficientNetB3 feature extraction, XAI using Grad-CAM, and classification using ERT and XG-Boost classifier.

Algorithm 1: ExHyptNet

Input: $X_{v,n}$ the number of PPG signals that hold (Normal tension, Hypertension, Pre-hypertension, v is the number of PPG signals in each dataset, while n is the number of samples.
Output: L the labels of the test features extracted from the ExHyptNet proposed model.
Stage1: RP Calculation
Step 1.1: Resize the $X_{v,n}$ to have the same number of samples in each PPG signal.
Step 1.2: Calculate the RP for each PPG signal in $X_{v,n}$
For k = 1 to v
$R_k = \Theta(\varepsilon - ||\overline{X}_{kt} - \overline{X}_{ku}||)t, u = 1, \cdots\cdots, N$

(continued on next page)

Table 4
Hyperparameters of EfficientNetB3 model.

Training parameters optimizer	Adam optimizer
Initial learning rate	0.001
Learning rate schedule	Piecewise
Learning rate drop factor	0.7
Learning rate drop period	10
Maximum number of epochs	50, 15
Mini-batch size	16
Gradient threshold method	L2 norm
Gradient threshold value	Inf
L2 regularization	1 x 10^{-3}
Verbose	1

Table 5
Parameters of the XGBoost and ERT Classifiers.

Parameters	XG-Boost	Parameters	ERT
Evaluation Matrix	Logistic Function	estimators	50
Booster	"gbtree" for multi-class "glinear" for binary class	Criterion or function of split	"log-loss"
Iteration	10	Iteration	15
Maximum Depth	5	Maximum Depth	10
Eta	0.1	*nmin* split	2.5
Gamma	0	*nmin* leaf	1.5

(*continued*)

Algorithm 1: ExHyptNet

Where k is the index that represents the current PPG signal, Θ is the Heaviside function,e is a predefined threshold nearly equal to zero,|||| is a norm, t and u are the time units in which any sample in the PPG signal is situated at times t and u.
End
Step 1.3: Obtain R for each PPG signal and as an input to the deep learning model.
Stage2: EfficientB3Net Feature Extraction
Step 2.1: Resize each RP in R_k to size of 224 x 224 x 3.
Step 2.2: Divide the RPs R into 60 % training RT, 20 % validation RV, and 20 % test RE images.
Step 2.3: Extract the features using the efficientNetB3 deep-learning model for the RT using the
following:
YT = EfficientNetB3(RT)
Step 2.4: Save the training model result in terms of net N.
Step 2.5: Get the features from the entire RPs R obtaining features Y.
Stage3: Explain the features using XAI
Step 3.1: Prepare the net N and all the RPs of the PPG signals R.
Step 3.2:
For k = 1 to m
XR_k-= Grad-CAM(R_k)
Where XR_k is the generated image of the RP after explaining the main important features in it using some colored spots on it.
End
Step 3.3: Save the generated images using the Grad-CAM XR_k.
Stage4: Classification
Step 4.1: Apply three types of cross-validation classification (Holdout, LOOCV, and SKCV) on the entire feature Y.
Step 4.2: Get the training features set Y_{train}, testing feature set Y_{test} from the features Y.
Step 4.3: Apply the ERT to the training data Y_{train} generating a subset node S
Start splitting the node S
If splitting == true
Select v attributes from S known as $\{a_1, a_2, a_3, \cdots\cdots a_v\}$ from node S.
Perform many v splits from $\{s_1, s_2, s_3, \cdots\cdots s_v\}$
s_v = Pick a random split (S, a_v), where i is the number of splits returned
Obtain the maximum score between max score (s_v, Y_{test}), where Y_{test} is the testing
data
Return L label corresponding to the maximum score reached using the ERT classifier
Else
The node S cannot split anymore
Return nothing
End
Step 4.4: Apply the XG-Boost on the training features Y_{train}, given the $Loss_{XGBoost}(Y_{test}, f(x))$, and the total number of sub-trees M, where $f(x)$ is the tree model of the XG-Boost classifier
Do
Initialize the $m-th$ tree $f_m(x_i)$
Calculate $g_i = \partial_{\hat{y}_i^{(m-1)}} Loss_{XGBoost}(y_i, \hat{y}_i^{(m-1)})$
Calculate $h_i = \partial^2_{\hat{y}_i^{(m-1)}} Loss_{XGBoost}(y_i, \hat{y}_i^{(m-1)})$
Where g and h are the first and the second derivatives of the loss function $Loss_{XGBoost}$, y_i and $\hat{y}$ are the true value and the estimated value of each feature vector in Y_{test}, and $Loss_{XGBoost}$ is the difference between the true and the estimated values
Start greedily growing the tree using the following objective function

$$obj^{(m)} = \frac{1}{2}\sum_{j=1}^{M}\frac{g_j^2}{h_j+\lambda} + \gamma M$$

Where λ is the penalty coefficient, and γ is a hyperparameter that controls the the complexity of the model

(*continued on next column*)

(*continued*)

Algorithm 1: ExHyptNet

Add the best tree $f_m(x_i)$ into the current model
While all the M are processed correctly
Get a strong classification tree relying on the weak classification sub-trees
Output the estimated probability of Y_{test} based on the produced strong classification
tree in terms of L

4.3. Classification results

Model performance was validated using three different validation strategies: (1) splitting the data into 60 % training and 40 % test sets after excluding the validation data; (2) LOOCV; and (3) stratified k-fold cross-validation (SKCV), with k = 10. Through internally splitting the data more than once, sampling bias was mitigated with the latter two approaches.

4.3.1. Classification results based on the PPG-BP dataset

The PPG-BP dataset comprised 657 PPG signals: 262 normal, 159 hypertension, and 236 prehypertension. The three Scenarios involved the classification of (1) normal vs. hypertension vs. prehypertension (training: validation: test data split 394:131:132); (2) normal vs. hypertension (training: validation: test data split 252:84:85); and (3) normal vs. prehypertension (training: validation: test data split 298:99:100). In all experiments, the number of epochs was set at 15.

Excellent results were obtained by splitting the data into 60 % training and 40 % test sets after excluding the validation data, LOOCV, and 10-fold CV (Table 6), with the ERT classifier attaining 100 % accuracies in all three Scenarios across all validation strategies. Confusion matrices in the respective cases demonstrate zero misclassification with the ERT classifier and low rates of misclassification with the XGBoost classifier (Figs. 3 to 5).

4.3.2. Classification results based on the MIMIC-II dataset

The MIMIC-II dataset comprised 120 PPG signals: 46 normal, 41 hypertension, and 34 prehypertension. The three experiments involved the classification of (1) normal vs. hypertension vs. prehypertension (training: validation: test data split 72:24:24); (2) normal vs. hypertension (training: validation: test data split 48:16:16); and (3) normal vs. prehypertension (training: validation: test data split 53:16:18). In all experiments, the number of epochs was set at 50.

Like with the PPG-BP dataset, perfect classification rates have been obtained with the ERT classifier by splitting the data into 60 % training and 40 % test sets after excluding the validation data, LOOCV, and 10-fold CV (Table 7), in all 3 experiments across all validation strategies, and the XGBoost classifier attaining 100 % accuracies in the second experiment, (normal vs. hypertension classification) with the 60 %: 40 % data split validation strategy. Confusion matrices in the respective cases demonstrate zero misclassification with the ERT classifier and low rates of misclassification with the XGBoost classifier (Figs. 6 to 8).

4.4. t-distributed Stochastic neighbor embedding (t-SNE)

We applied t-SNE (Van der Maaten & Hinton, 2008) on both raw PPG signals and the feature maps extracted by EfficientNetB3, to perform the unsupervised randomized reduction of the high-dimensional data to lower dimensional spaces (two-dimensional graphs in our case), facilitate visualization of nonlinear clustering of similar data points using parameter-base skewed Student's t probability distributions (Huang et al., 2019). For both PPG-BP and MIMIC-II datasets, the t-SNE plots of the raw PPG signals appear scattered before feature extraction. In contrast, the extracted feature maps show separated clusters corresponding to the normal, hypertension, and prehypertension groups (Fig. 9). Assuming that neighboring points in the low-dimensional t-SNE plots are also likely to be neighboring points in the input high-

Table 6
Model performance on the PPG-BG dataset stratified by validation strategies, classification categories, and classifiers.

Classifier	TP	FP	Pre	Rec	F1	MMC	ROC	PRC	κ	Acc (%)
60 % TRAINING AND 40 % TEST DATA SPLIT										
	Normal vs. Prehypertension vs. Hypertension									
XGBoost	0.99	0.01	0.99	0.99	0.99	0.99	0.99	0.99	0.99	99.24
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Hypertension									
XGBoost	0.95	0.08	0.96	0.95	0.95	0.90	0.94	0.93	0.90	95.29
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Prehypertension									
XGBoost	0.95	0.06	0.95	0.95	0.95	0.90	0.95	0.93	0.90	95.00
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
LEAVE-ONE-OUT CROSS-VALIDATION										
	Normal vs. Prehypertension vs. Hypertension									
XGBoost	1.00	0.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99	99.64
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Hypertension									
XGBoost	0.99	0.01	0.99	0.99	0.99	0.99	1.00	0.99	0.99	99.43
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Prehypertension									
XGBoost	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	99.76
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
10-FOLD CROSS-VALIDATION										
	Normal vs. Prehypertension vs. Hypertension									
XGBoost	1.00	0.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99	99.64
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Hypertension									
XGBoost	1.00	0.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	99.72
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Prehypertension									
XGBoost	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	99.76
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00

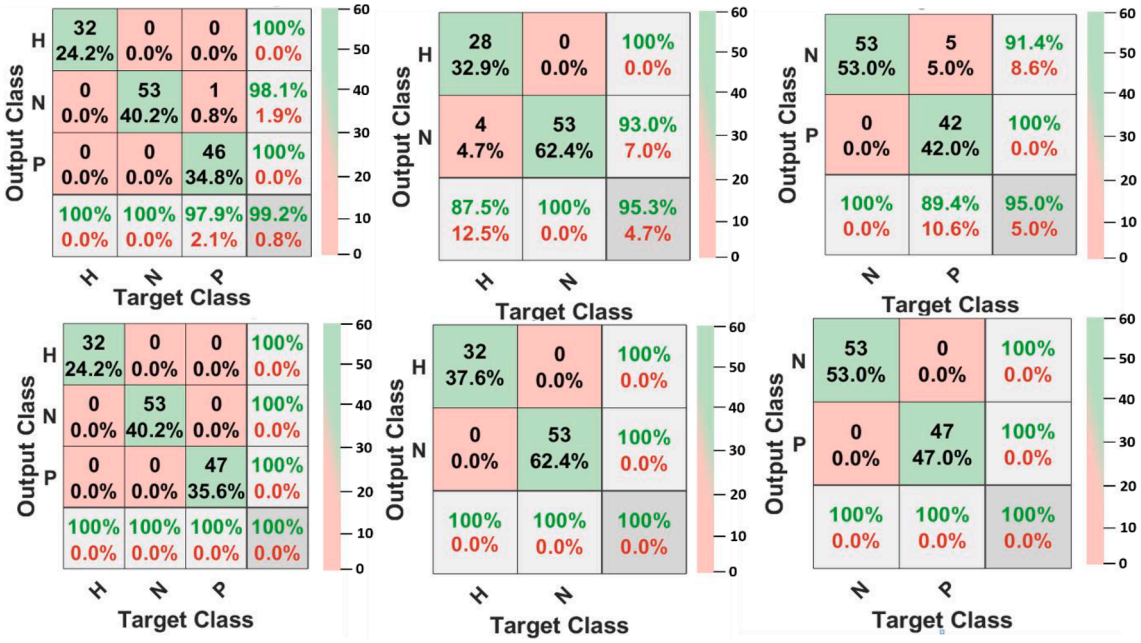

Fig. 3. Confusion matrices were obtained for 3 experiments using PPG-BP dataset with XGBoost classifier (top row) and ERT classifier (bottom row). H, hypertension; N, normal; P, prehypertension. In this work, we used 60% training and 40% test data split.

dimensional space, we could surmise that the feature maps extracted by EfficientNetB3 from the RPS were discriminative for the different classes in both datasets.

4.5. Results of explainable-AI using Grad-CAM

Grad-CAM allows visualization of the gradients of the classification scores relative to the final convolutional feature maps. The parts of the map with high values on the Grad-CAM map signify those with the most impact on the network score for that class. By differentiating the reduced output of the reduction layer concerning the features in the feature layer, Grad-CAM automatically selects the reduction and feature layers to use when computing the map. In our study, Grad-CAM readouts obtained after PPG RPs trained by the EfficientNetB3 model were back propagated on the RPs as heat maps highlighting areas of the plots that contributed the most to the network classification decisions for the

Fig. 4. Confusion matrices were obtained for 3 experiments using PPG-BP dataset with XGBoost classifier (top row) and ERT classifier (bottom row). H, hypertension; N, normal; P, prehypertension. In this work, we employed a leave-one-out cross-validation strategy.

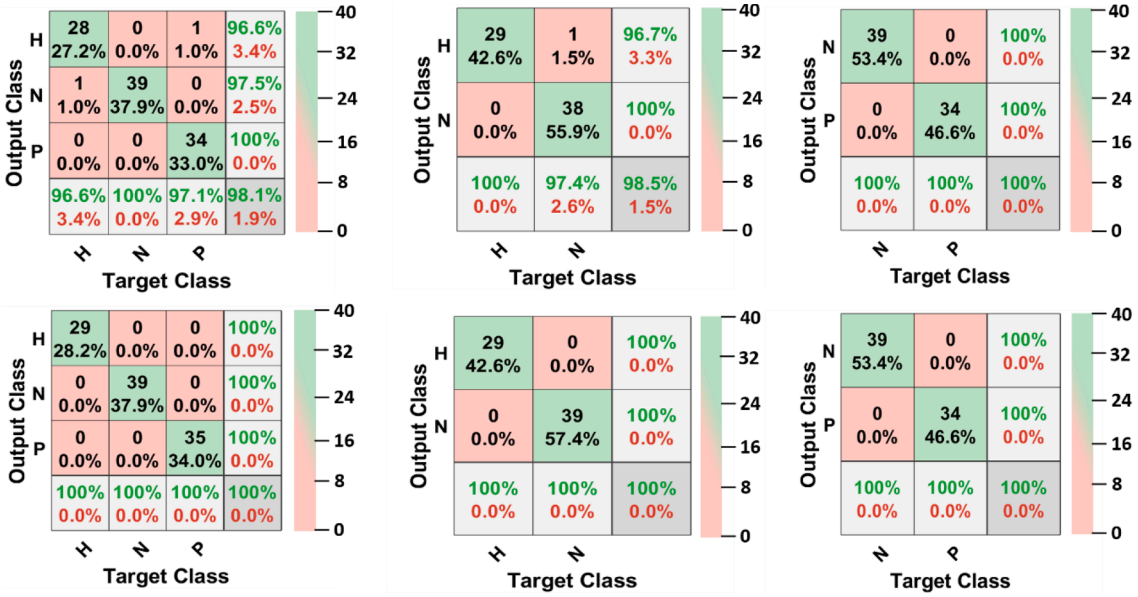

Fig. 5. Confusion matrices were obtained for 3 experiments using the PPG-BP dataset with XGBoost classifier (top row) and ERT classifier (bottom row). H, hypertension; N, normal; P, prehypertension. In this work, we used a 10-fold cross-validation strategy.

individual classes. The RP and the Grad-CAM have provided vital information that could provide assistance and guidance to the physicians. The vertical and horizontal planes in the RP images are the key results of hypertension detection. It can be seen how they indicate slowly changing or completely steady states. The Grad-DCAM allows researchers and clinical experts to identify potential regions in the RP responsible for detecting HC or HYPT. Therefore, our model helps to detect and identify the portion of the RP which contributes to the decision-making. In the future, we will include clinical features and explore model explainability on these features. It will help the clinical experts to detect the potential reasons for hypertension.

The coloured regions presented in the RP obtained from the Grad-CAM are due to the trained network features obtained from the EfficientNetB3 on the RP images. Grad-CAM contributed to visualizing and converting these solid features into several coloured regions with different shapes on the RP images. It can be visualized that the coloured regions provided are the vital features and components in the RP images that helped the classifier to differentiate between different types of hypertension. It is important to mention how the Grad-CAM chooses or draws these regions on the RP to provide a perfect explanation of the proposed model. The Grad-CAM method employs the gradients of the classification score concerning the last convolutional feature map. This is done to pinpoint the coloured regions within an input image (RP) that have the greatest influence on the classification score. These regions with significant gradient values precisely indicate where the classification outcome is most influenced by the data.

Fig. 10 shows the RP images in the left column and the Grad-CAM obtained from three RP images representing (normal, prehypertension, and hypertension) in the right column.

Each image in the right column has three different regions (bright,

Table 7
Model performance obtained using the MIMIC-II dataset for various validation strategies, classification categories, and classifiers.

Classifier	TP	FP	Pre	Rec	F1	MMC	ROC	PRC	κ	Acc (%)
60 % TRAINING AND 40 % TEST DATA SPLIT										
	Normal vs. Prehypertension vs. Hypertension									
XGBoost	0.96	0.02	0.96	0.96	0.96	0.94	1.00	1.00	0.94	96.00
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Hypertension									
XGBoost	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Prehypertension									
XGBoost	0.89	0.09	0.91	0.89	0.89	0.80	1.00	1.00	0.78	88.88
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
LEAVE-ONE-OUT CROSS-VALIDATION										
	Normal vs. Prehypertension vs. Hypertension									
XGBoost	0.95	0.02	0.95	0.95	0.95	0.93	0.98	0.96	0.93	95.14
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Hypertension									
XGBoost	0.99	0.01	0.99	0.99	0.99	0.97	0.99	0.98	0.97	98.53
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Prehypertension									
XGBoost	0.97	0.02	0.97	0.97	0.97	0.95	1.00	1.00	0.95	97.26
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
10-FOLD CROSS-VALIDATION										
	Normal vs. Prehypertension vs. Hypertension									
XGBoost	0.98	0.01	0.98	0.98	0.98	0.97	0.99	0.97	0.97	98.06
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Hypertension									
XGBoost	0.99	0.01	0.99	0.99	0.99	0.97	0.99	0.98	0.97	98.53
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	Normal vs. Prehypertension									
XGBoost	0.99	0.02	0.99	0.99	0.99	0.97	0.99	0.98	0.97	98.63
ERT	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00

Fig. 6. Confusion matrices were obtained for 3 experiments using MIMIC-II dataset with XGBoost classifier (top row) and ERT classifier (bottom row). H, hypertension; N, normal; P, prehypertension. In this work, we used 60% training and 40% test data split.

medium dark, and dark). The dark and the medium dark regions represent the most significant and salient parts in the RP images that affected the classification performance, while the bright regions show the less important feature parts in the RP images.

5. Ablation study

We have conducted an ablation study on the block of the pre-trained model and by increasing the number of categories on the PPG-BP dataset.

Hyperparameters: Several values are assigned to the hyperparameters for the EfficientNetB3 to obtain the optimal values for the

Fig. 7. Confusion matrices were obtained for 3 experiments using MIMIC-II dataset with XGBoost classifier (top row) and ERT classifier (bottom row). H, hypertension; N, normal; P, prehypertension. In this work, we used a leave-one-out cross-validation strategy.

Fig. 8. Confusion matrices were obtained for 3 experiments using the MIMIC-II dataset with XGBoost classifier (top row) and ERT classifier (bottom row). H, hypertension; N, normal; P, prehypertension. In this work, we used a 10-fold cross-validation strategy.

deep learning model. As mentioned before, three different experiments are performed on each dataset. Table 8 shows the values assigned to the hyperparameters.

Different pre-trained model: The current pre-trained model used in the proposed study is the EfficientNetB3 on both datasets. Several versions of EfficientNet exist, therefore EfficientNetB4 was tried on several experiments using two datasets. It was discovered that the EfficientNetB4 with ERT classifier achieved perfect performance using PPG-BP and MIMIC-II datasets. Table 9 shows the performance of the ERT classifier using EfficientNetB4 with holdout data using PPG-BP and MIMIC-II datasets.

Finally, it is important to mention that EfficientNetB4 is more complex than EfficientNetB3. EfficientNetB3 has approximately 12 million parameters, while EfficientNetB4 has around 19 million parameters (Koonce & Koonce, 2021). The increase in model complexity allows EfficientNetB4 to capture more intricate features and potentially achieve better performance on certain tasks (Zhang et al., 2020). However, it also comes with increased computational requirements during training and inference. The training time required using EfficientNetB4 on the PPG-BP dataset is 50 mins and 1 h &15 min for binary and multi-classes, respectively. Similarly, the training time using EfficientNet B3 on the PPG-BP dataset is 16 mins and 37 mins for binary and multi-classes, respectively.

Increasing the number of classes: Three experiments are performed using two datasets. These experiments are (i) (Normal vs Hypertension vs prehypertension), (ii) (Normal vs Hypertension) and (iii) (Normal vs Prehypertension). In the multi-class experiment, the hypertension stage classes in the PPG-BP are divided into two main stages

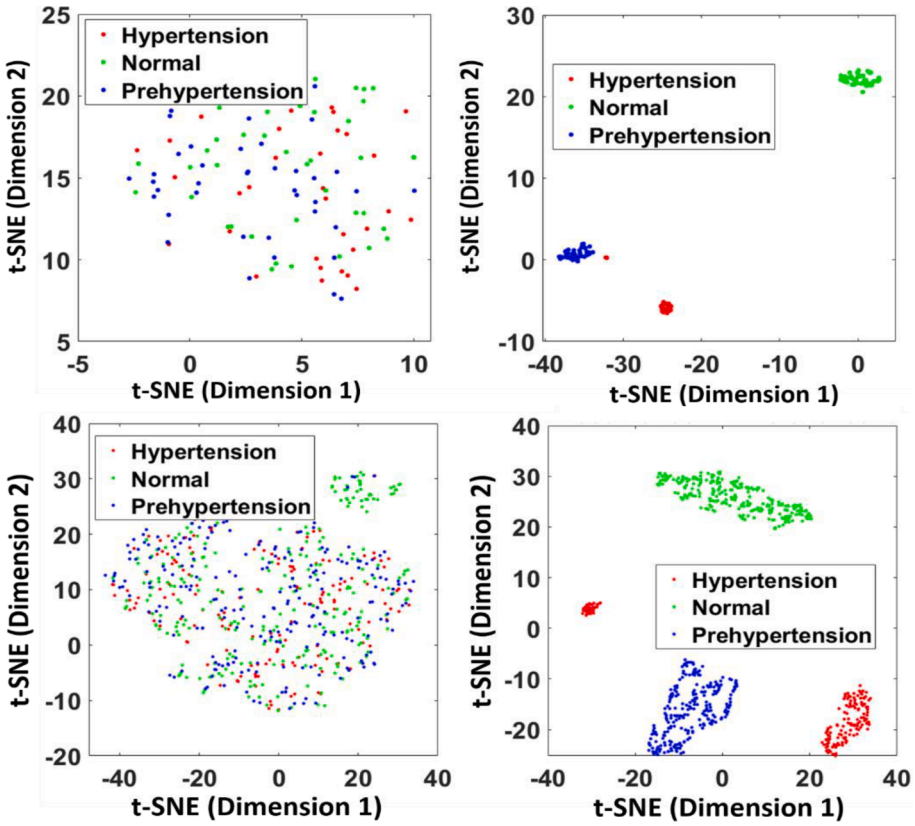

Fig. 9. t-SNE of input PPG signals (left column) and EfficientNetB3 output feature maps (right column) from the PPG-BP (top row) and MIMIC-II (bottom row) dataset. The axes in the plots hold no particular meaning.

(Stage 1 Hypertension and Stage 2 Hypertension). Therefore, it is decided to split the data and perform the experiment based on (Normal, vs Prehypertension vs Stage 1 hypertension vs Stage 2 hypertension) using the proposed methodology with ERT classifiers based on 10 fold cross-validation strategy. The final accuracy of 98 % is obtained for 4 classes using the PPG-BP dataset with the ERT classifier. Table. 10 shows the model performance after dividing the hypertension into two stages (Stage1 Hypertension and Stage 2 Hypertension).

6. Discussion

Our ExHyptNet model outperformed existing hypertension classification models developed using PPG-BP and MIMIC-II datasets (Table 11). It is evident from Table 11 that most of the models developed previously for hypertension detection have used a single dataset to test their model. In addition, the models have been developed either using k-fold cross-validation or holdout validation techniques. The highest accuracy obtained on the PPG-BP dataset has been reported to be 99.5 % using fivefold and k-fold cross-validation (Nour & Polat, 2020; Sadad et al., 2022). Similarly, the highest accuracy obtained on the MIMIC-II dataset has been reported to be 99.4 % using the holdout validation technique (Sannino et al., 2020). In addition, only one study has used both datasets to validate their model performance and obtained an accuracy of 69.9 % and 72.9 % on the PPG-BP and MIMIC-II datasets, respectively (Evdochim et al., 2022). The models developed using holdout validation techniques are susceptible to overfitting due to the bias in the train and test split.

Similarly, the models developed using k-fold cross-validation may not find real-time clinical acceptance. Also, there can be scenarios, where a model developed on one dataset with a fixed system setting may not produce the desired performance if the same model is tested on another dataset with a different system or setting. Similarly, the existing models did not discuss the model explainability, which clinicians may find difficult to accept the high performance of the model (Khare & Acharya, 2023). Our developed model addresses these shortcomings as it has been tested on two public datasets using different validation techniques and explores model explainability. Our developed model has achieved a perfect classification rate with an Acc of 100 % using holdout, SKCV, and LOOCV techniques on both datasets.

The combination of RP images on PPG signals and EfficientNetB3 has resulted in a robust feature extraction. RPs are a powerful tool for several reasons. Firstly, they can analyze the periodicity of the PPG signal. Secondly, they can deal with the non-stationarity present in the PPG signals. Thirdly, they can identify any hidden patterns and irregularities in the PPG signals. The usage of EfficientNetB3 is advantageous for two main reasons. Firstly, it can achieve high accuracy compared to other pre-trained models. Secondly, it can achieve fast throughput and meet the requirements of a quick response in the diagnosis of hypertension.

It is worth mentioning that the implementation of Grad-CAM was vital for three main reasons. Firstly, it can generate a visual explanation that makes the combination of the RP and EfficientNetB3 more transparent. Secondly, it produces a location map that highlights the main regions on the RP image that were affected by the classifiers. Thirdly, it can guide the system in the direction of learning the model better.

The proposed model performed three main experiments on both PPG-BP and MIMIC-II datasets. These experiments included multi-class (Hypertension vs Normal vs Prehypertension) and two binary-class (Hypertension vs Normal) and (Hypertension vs Prehypertension), and they were applied to explore the efficiency of the extracted features. Additionally, two main classifiers (ERT and XG-Boost) were applied to validate the performance of the proposed model. It was observed that ERT outperformed XG-Boost in all experimental scenarios.

The complexity of the ExHyptNet model is influenced by multiple

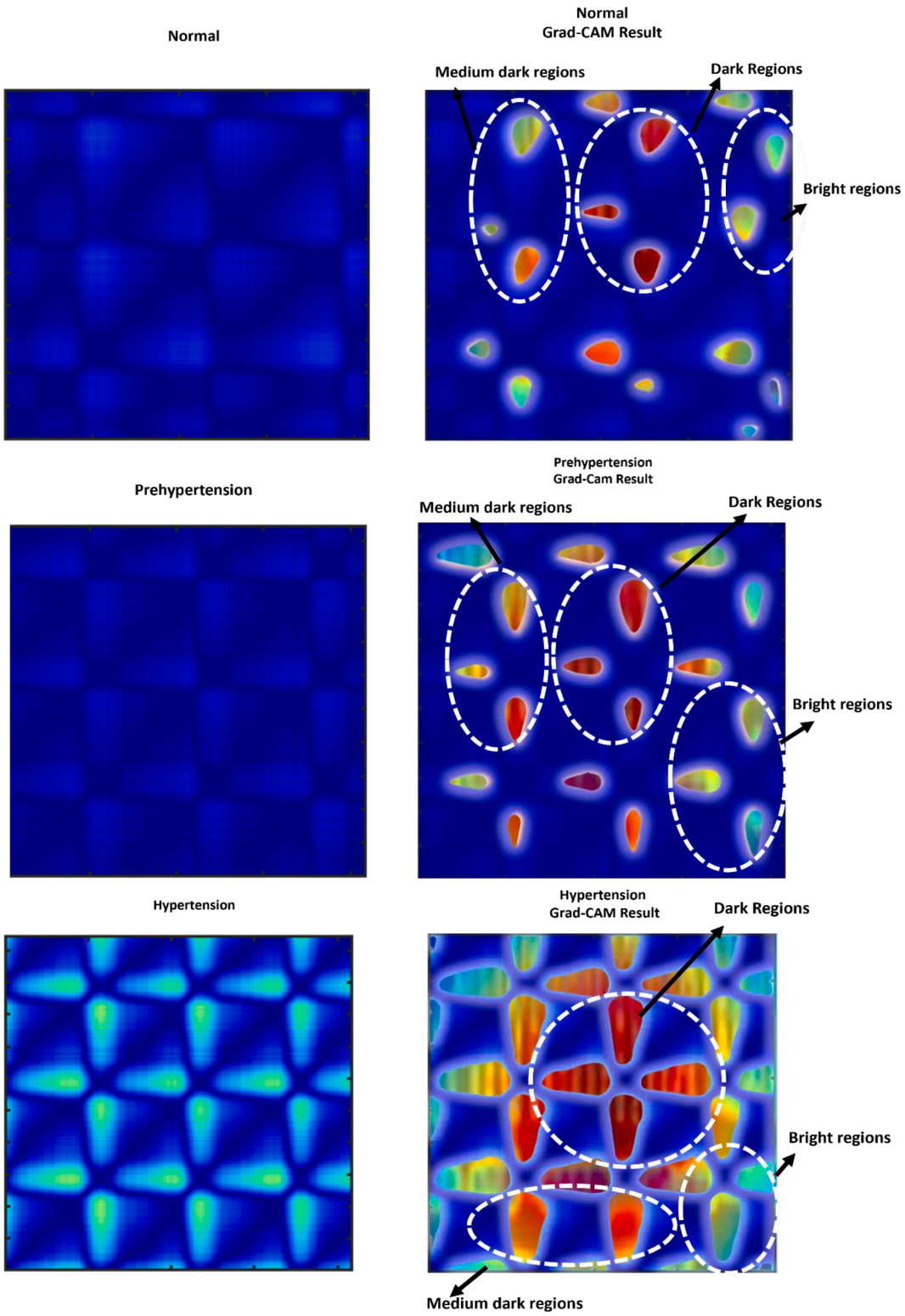

Fig. 10. Example of RPs (left column) in the normal (top row), prehypertension (middle row), and hypertension (bottom row) classes. The same plots are overlayed with Grad-CAM heat maps generated after training with EfficientNetB3 (right column) using the ERT classifier. The Grad-CAM enables visualization of parts of the RPs that exert the most impact on the network classification decision for each class.

elements, including the duration of learning, data volume, model layers, and the quantity and dimensions of filters within each layer. The assessment of the model's intricacy is conducted by estimating learning duration, deriving features from the fully connected layers, and categorizing the evaluation data. The training time of the ExHyptNet model on the three experiments on the PPG-BP dataset is 17.2, 13.7, and 12.3 min, while the training time of the ExHyptNet model on the three experiments on the MIMIC-II dataset is 32.3, 23.2, and 24.4 min. Most of the learning time is spent on training the EfficientNetB3 model, while a little time is spent on the Grad-CAM, ERT, and XG-Boost classifiers. The time necessary to obtain the RP images from the entire PPG-BP dataset is 5.6 min, while the time for the MIMIC-II dataset is 15.2 min. The time

Table 8
Hyperparameter values assigned to the EfficientNetB3 pre-trained model.

Training parameters optimizer	Adam	Adam	Adam	Adam	Adam
Initial learning rate	0.001	0.0001	0.005	0.0003	0.002
Learning rate drop factor	0.7	0.4	0.8	0.5	0.2
Learning rate drop period	10	12	19	8	4
Mini-batch size	16	8	32	30	64
L2 regularization	1 x 10^{-3}	1 x 10^{-5}	1 x 10^{-6}	1 x 10^{-2}	1 x 10^{-4}
Validation Accuracy	99.5 %	97.7 %	98.2 %	97.9 %	98.3 %

required to extract features from the EfficientNetB3 model is 5.2 min. The time taken to classify the features from the Vgg16-GRU model in the experiment is 2.1 min.

The advantages of our developed model are listed below:

i. Versatile: The ExHyptNet model has been developed using two public PPG datasets with three validation techniques. Hence the generated system is versatile for different scenarios.
ii. Accurate: Our model has yielded the highest multi- and binary-class classification performance without misclassification.
iii. Reliable: The ExHyptNet model highlights the region in RP responsible for hypertension detection, making it reliable.
iv. Robust: The developed model has proven effective for qualitative and quantitative analysis, making it robust for data-balanced and unbalanced scenarios.
v. XAI: Proposed unique Grad-CAM heat maps for each class to develop trust in researchers and clinicians. In this work, we have obtained unique heatmaps for normal, prehypertension, and hypertension classes.
vi. To the best of our knowledge it is the first work to use two public databases and obtain 100 % classification performance with XAI.

The limitations of our model are:

Table 9
Model performance obtained using MIMIC-II and PPG-BP datasets using the proposed methodology of EfficientNet B4 coupled with ERT classifier.

	TP	FP	Pre	Rec	F1	MMC	ROC	PRC	κ	Acc (%)
60 % TRAINING AND 40 % TEST DATA SPLIT										
	Normal vs. Prehypertension vs. Hypertension									
MIMIC-II										
PPG-BP	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
Normal vs. Hypertension										
MIMIC-II	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
PPG-BP	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
Normal vs. Prehypertension										
MIMIC-II	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00
PPG-BP	1.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	100.00

Table 10
Model performance obtained using our proposed method with ERT classifier for the PPG-BP dataset.

	TP	FP	Pre	Rec	F1	MMC	ROC	PRC	κ	Acc (%)
10-FOLD CROSS-VALIDATION										
	Normal vs. Prehypertension vs. Stage 1 Hypertension vs Stage 2 Hypertension									
PPG-BP	0.98	0.01	0.98	0.98	0.98	0.98	1.00	1.00	0.98	98.00

Table 11
Comparison of our model with state-of-the-art methods developed for hypertension classification using the PPG-BP and MIMIC-II datasets.

Study	Dataset	Subjects	Features	Validation	Classifier	Acc (%)	Rec (%)	Spe (%)
Mousavi et al., 2019	MIMIC-II	441 Subjects 1323 records	Whole-based features	10-fold CV	SVM	–	–	–
Nour & Polat, 2020	PPG-BP	219 Subjects 657 records	Physiological features	5-fold CV	RF	99.5	–	–
Evdochimet al., 2022	PPG-BP,	219 Subjects	Signal morphology	–	Decision tree	69.9	66.7	71.1
	MIMIC-II	140 Subjects			SVM	72.9	71	67.5
Gupta et al., 2022	PPG-BP	219 Subjects	Filtering	–	CNN-LSTM	67.76	68.4	66.6
Khan et al. (2022)	PPG-BP	219 Subjects 657 records	VMD	10-fold CV	Gradient boosting	99.3	98.7	100
Ranjan et al., 2022	PPG-BP	219 Subjects 657 records	Filtering	–	CNN	74.5	73.4	75.7
Sadadet al., 2022	PPG-BP	219 Subjects 657 records	Heart rate	k-fold CV	Decision tree	99.5	–	–
Tanc and Ozturk (2022)	MIMIC-II	150 records	SST	Holdout	GoogLeNet	–	95.78	95.96
Yen et al., 2021	PPG-BP	219 subjects 657 records	–	Holdout	ResNetCNN	73	–	–
Mejía-Mejía et al. (2021)	MIMIC-II	32,536 Subjects	Nonlinear features	10-fold CV	SVM	70	50	75
Sannino et al., 2020	MIMIC-II	526,906 instances	–	Holdout	RF	99.4	–	–
Our study	PPG-BP,	219 Subjects 657 records	RPs, EfficientNetB3	Holdout	ERT	100	100	100
	MIMIC-II	120 Subjects		10-fold CV		100	100	100
				LOOCV		100	100	100

Spe-Specificity; CV- cross-validation; CNN– convolutional neural network; LSTM- long short-term memory; RF- random forest; SVM- support vector machine; VMD- variational mode decomposition; SST- Synchrosqueezing transform.

i. Offline system: The model is developed for the offline system; hence, the effects of online training are unknown.
ii. Our developed model is generated in a noise-free environment. But in the real-world scenario, the generated model may not yield the same high performance.
iii. Dataset: The proposed model has been generated using two small public datasets. More diverse datasets obtained from various centres need to be used.

In the future, we plan to validate our model using the dataset obtained from many centres and also evaluate the performance of the model in a noisy environment using uncertainty quantification (UQ) in concept (Seoni et al., 2023).

7. Conclusion

In summary, the timely detection of hypertension is critical for preventing heart-related problems. Our ExHypNet model narrows down the gaps in the current PPG-based automated hypertension detection systems. Our model has obtained the highest performance on two public PPG datasets using different validation strategies. The quantitative and qualitative analysis shows that our developed model is accurate, effective, and versatile as it is tested on two independent datasets using multi- and binary-class classification for both balanced and unbalanced data distributions. To the best of our knowledge, we are the first to develop a versatile and easily understandable hypertension detection system for offline use. Our developed ExHypNet model is ready to be deployed in the clinical setting for real-time hypertension detection. Furthermore, our ablation study demonstrates that using EfficientB3Net and EfficientB4Net with the ERT classifier results in 100 % accuracy. However, we propose using EfficientB3Net due to its simpler structure, fewer learning parameters, and faster feature extraction, especially when classifying four different categories. In the future, our research will explore uncertainty quantification in diverse noise scenarios and aim to develop an online classification model to effectively detect various cardiovascular diseases.

CRediT authorship contribution statement

El-Sayed A. El-Dahshan: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Mahmoud M. Bassiouni:** Methodology, Writing – original draft, Writing – review & editing. **Smith K. Khare:** Methodology, Writing – original draft, Writing – review & editing, Visualization. **Ru-San Tan:** Methodology, Writing – original draft, Writing – review & editing, Visualization. **U. Rajendra Acharya:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Visualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors are unable or have chosen not to specify which data has been used.

References

Bassiouni, M. M., El-Dahshan, E. S. A., Khalefa, W., & Salem, A. M. (2018). Intelligent hybrid approaches for human ECG signals identification. *Signal, Image and Video Processing, 12*, 941–949.

Loh, H. W., Xu, S., Faust, O., Ooi, C. P., Barua, P. D., Chakraborty, S., & Acharya, U. R. (2022). Application of photoplethysmography signals for healthcare systems: An in-depth review. In *Computer Methods and Programs in Biomedicine* (p. 106677).

Organization, W. H. (2013). *A global brief on hypertension: Silent killer, global public health crisis: World Health Day 2013*. World Health Organization.

Al-Zaben, A., Fora, M., & Obaidat, A. (2018). In *Detection of premature ventricular beats from arterial blood pressure signal* (pp. 17–19). IEEE.

El-Dahshan, E. S. A., & Bassiouni, M. M. (2018). Computational intelligence techniques for human brain MRI classification. *International Journal of Imaging Systems and Technology, 28*(2), 132–148.

Hall, J. E. (2003). The kidney, hypertension, and obesity. *Hypertension, 41*(3), 625–633.

Drozdz, D., & Kawecka-Jaszcz, K. (2014). Cardiovascular changes during chronic hypertensive states. *Pediatric Nephrology, 29*, 1507–1516.

Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Informatics in Medicine Unlocked, 17*, Article 100179.

Azadifar, S., Rostami, M., Berahmand, K., Moradi, P., & Oussalah, M. (2022). Graph-based relevancy-redundancy gene selection method for cancer diagnosis. *Computers in Biology and Medicine, 147*, Article 105766.

Maqsood, S., Xu, S., Springer, M., & Mohawesh, R. (2021). A benchmark study of machine learning for analysis of signal feature extraction techniques for blood pressure estimation using photoplethysmography (PPG). *IEEE Access, 9*, 138817–138833.

Almarshad, M. A., Islam, M. S., Al-Ahmadi, S., & BaHammam, A. S. (2022, March). Diagnostic Features and Potential Applications of PPG Signal in Healthcare: A Systematic Review. In *Healthcare* (Vol. 10, No. 3, p. 547). MDPI.

Scheer, B. V., Perel, A., & Pfeiffer, U. J. (2002). Clinical review: Complications and risk factors of peripheral arterial catheters used for haemodynamic monitoring in anaesthesia and intensive care medicine. *Critical Care, 6*, 1–7.

Tjahjadi, H., & Ramli, K. (2020). Noninvasive blood pressure classification based on photoplethysmography using k-nearest neighbors algorithm: A feasibility study. *Information, 11*(2), 93.

Hettiarachchi, C., & Chitraranjan, C. (2019). In *A machine learning approach to predict diabetes using short recorded photoplethysmography and physiological characteristics* (pp. 322–327). Springer International Publishing.

Mousavi, S. S., Firouzmand, M., Charmi, M., Hemmati, M., Moghadam, M., & Ghorbani, Y. (2019). Blood pressure estimation from appropriate and inappropriate PPG signals using A whole-based method. *Biomedical Signal Processing and Control, 47*, 196–206.

Yao, L. P., & Liu, W. Z. (2021). Hypertension assessment based on feature extraction using a photoplethysmography signal and its derivatives. *Physiological Measurement, 42*(6), Article 065001.

Cano, J., Bertomeu-González, V., Fácila, L., Zangróniz, R., Alcaraz, R., & Rieta, J. J. (2021). *September). Hypertension Risk Assessment from Photoplethysmographic Recordings Using Deep Learning Classifiers* (Vol. 48,, 1–4.

Nour, M., & Polat, K. (2020). Automatic classification of hypertension types based on personal features by machine learning algorithms. *Mathematical Problems in Engineering, 2020*, 1–13.

Khan, M. U., Aziz, S., Akram, T., Amjad, F., Iqtidar, K., Nam, Y., & Khan, M. A. (2021). Expert hypertension detection system featuring pulse plethysmograph signals and hybrid feature selection and reduction scheme. *Sensors, 21*(1), 247.

Evdochim, L., Dobrescu, D., Halichidis, S., Dobrescu, L., & Stanciu, S. (2022). Hypertension Detection Based on Photoplethysmography Signal Morphology and Machine Learning Techniques. *Applied Sciences, 12*(16), 8380.

Gupta, K., Jiwani, N., & Afreen, N. (2022 April). In *Blood Pressure Detection Using CNN-LSTM Model* (pp. 262–366). IEEE.

Khan, M., Singh, B. K., Nirala, N., Hasan, M. T., & Nasiruddin, M. (2022). Multi-Domain Feature-based Expert Diagnostic System for Detection of Hypertension using Photoplethysmogram Signal. *International Journal of Intelligent Systems and Applications in Engineering, 10*(4), 424–433.

Ranjan, P., Kumar, A., Kumar, R., Sameer, M., & Gupta, B. (2022). In *February). Automated Detection of Blood Pressure using CNN* (pp. 1–5). IEEE.

Sadad, T., Bukhari, S. A. C., Munir, A., Ghani, A., El-Sherbeeny, A. M., & Rauf, H. T. (2022). Detection of Cardiovascular Disease Based on PPG Signals Using Machine Learning with Cloud Computing. *Computational Intelligence and Neuroscience, 2022*.

Tanc, Y. H., & Ozturk, M. (2022). In *October). Hypertension Classification Using PPG Signals* (pp. 1–4). IEEE.

Gupta, S., Singh, A., Sharma, A., & Tripathy, R. K. (2022). Higher Order Derivative-Based Integrated Model for Cuff-Less Blood Pressure Estimation and Stratification Using PPG Signals. *IEEE Sensors Journal, 22*(22), 22030–22039.

Wu, J., Liang, H., Ding, C., Huang, X., Huang, J., & Peng, Q. (2021). Improving the accuracy in classification of blood pressure from photoplethysmography using continuous wavelet transform and deep learning. *International Journal of Hypertension, 2021*.

Liang, Y., Chen, Z., Ward, R., & Elgendi, M. (2018). Photoplethysmography and deep learning: Enhancing hypertension risk stratification. *Biosensors, 8*(4), 101.

Yen, C. T., Chang, S. N., & Liao, C. H. (2021). Deep learning algorithm evaluation of hypertension classification in less photoplethysmography signals conditions. *Measurement and Control, 54*(3–4), 439–445.

Nafisi, V. R., & Shahabi, M. (2018). Intradialytic hypotension related episodes identification based on the most effective features of photoplethysmography signal. *Computer Methods and Programs in Biomedicine, 157*, 1–9.

Mejía-Mejía, E., May, J. M., Elgendi, M., & Kyriacou, P. A. (2021). Classification of blood pressure in critically ill patients using photoplethysmography and machine learning. *Computer Methods and Programs in Biomedicine, 208*, Article 106222.

Lee, S., Lee, H. C., Chu, Y. S., Song, S. W., Ahn, G. J., Lee, H., & Koh, S. B. (2021). Deep learning models for the prediction of intraoperative hypotension. *British Journal of Anaesthesia, 126*(4), 808–817.

Ardiny, H., Witwicki, S., & Mondada, F. (2015, October). Construction automation with autonomous mobile robots: A review. In *2015 3rd RSI International Conference on Robotics and Mechatronics (ICROM)* (pp. 418-424). IEEE.

Sannino, G., De Falco, I., & De Pietro, G. (2020, December). Photoplethysmography and machine learning for the hypertension risk stratification. In *2020 IEEE Globecom Workshops (GC Wkshps* (pp. 1-6). IEEE.

Lan, K. C., Raknim, P., Kao, W. F., & Huang, J. H. (2018). Toward hypertension prediction based on PPG-derived HRV signals: A feasibility study. *Journal of Medical Systems, 42*, 1–7. https://doi.org/10.1007/s10916-018-0942-5

Marwan, N., Kurths, J., & Saparin, P. (2007). Generalised recurrence plot analysis for spatial data. *Physics Letters A, 360*(4–5), 545–551.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Zhou, T. (2015). *Xgboost: extreme gradient boosting., 1*(4), 1–4. R package version 0.4-2,.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). *Extremely randomized trees. Machine learning, 63*, 3–42.

Liang, Y., Chen, Z., Liu, G., & Elgendi, M. (2018). A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. *Scientific data, 5*(1), 1–7.

Lee, J., Scott, D. J., Villarroel, M., Clifford, G. D., Saeed, M., & Mark, R. G. (2011, August). Open-access MIMIC-II database for intensive care research. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 8315-8318). IEEE.

Lee, H., Lee, J., & Shin, M. (2019). Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots. *Electronics, 8*(2), 192.

Roh, D., & Shin, H. (2021). Recurrence plot and machine learning for signal quality assessment of photoplethysmogram in mobile environment. *Sensors, 21*(6), 2188.

Jan, H. Y., Chen, M. F., Fu, T. C., Lin, W. C., Tsai, C. L., & Lin, K. P. (2019). Evaluation of coherence between ECG and PPG derived parameters on heart rate variability and respiration in healthy volunteers with/without controlled breathing. *Journal of Medical and Biological Engineering, 39*, 783–795.

Barroso-García, V., Gutiérrez-Tobal, G. C., Kheirandish-Gozal, L., Álvarez, D., Vaquerizo-Villar, F., Núñez, P., & Hornero, R. (2020). Usefulness of recurrence plots from airflow recordings to aid in paediatric sleep apnoea diagnosis. *Computer Methods and Programs in Biomedicine, 183*, Article 105083.

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

Kashevnik, A., & Ali, A. (2022). 3D Vehicle Detection and Segmentation Based on EfficientNetB3 and CenterNet Residual Blocks. *Sensors, 22*(20), 7990.

Ganesh, M., Dulam, S., & Venkatasubbu, P. (2021). Diabetic Retinopathy Diagnosis with InceptionResNetV2, Xception, and EfficientNetB3. In *Artificial Intelligence and Technologies: Select Proceedings of ICRTAC-AIT 2020* (pp. 405-413). Singapore: Springer Singapore.

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: An analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11*(5), e1424.

Selvaraju, R. R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., & Batra, D. (2016). Grad-CAM: Why did you say that?. *arXiv preprint arXiv:1611.07450*.

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine, 107161*.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 1189–1232.

Saeed, U., Jan, S. U., Lee, Y. D., & Koo, I. (2021). Fault diagnosis based on extremely randomized trees in wireless sensor networks. *Reliability Engineering & System Safety, 205*, Article 107284.

Sharaff, A., & Gupta, H. (2019). Extra-tree classifier with metaheuristics approach for email classification. In *Advances in Computer Communication and Computational Sciences: Proceedings of IC4S 2018* (pp. 189-197). Springer Singapore.

García, S., Fernández, A., Luengo, J., & Herrera, F. (2009). A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. *Soft Computing, 13*, 959–977.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*(11).

Huang, Y., Zhang, Y., Zhao, Y., & Chambers, J. A. (2019). A novel robust Gaussian–Student's t mixture distribution based Kalman filter. *IEEE Transactions on signal Processing, 67*(13), 3606–3620.

Khare, S. K., & Acharya, U. R. (2023). An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals. *Computers in Biology and Medicine, 155*, Article 106676.

Seoni, S., Jahmunah, V., Salvi, M., Barua, P. D., Molinari, F., & Acharya, U. R. (2023). Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013–2023). *Computers in Biology and Medicine, 107441*.

Koonce, B., & Koonce, B. (2021). *EfficientNet* (pp. 109–123). Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization.

Zhang, P., Yang, L., & Li, D. (2020). EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment. *Computers and Electronics in Agriculture, 176*, Article 105652.