

Supplementary Material

Classification into covered or uncovered patients achieved an accuracy of 95.52%. Since the patients are kept under observations for long hours in intensive care conditions, it is possible for them to remain wrapped in a blanket for most of the time. Using ResNet-152^[13], we were able to classify the video frames into covered/uncovered to get a hold of any window during which the blanket is removed. An accuracy of 95.52% was achieved, with an F1 score of 0.9589 (Precision=0.9722, Recall=0.9459), creating less chances of false positives.

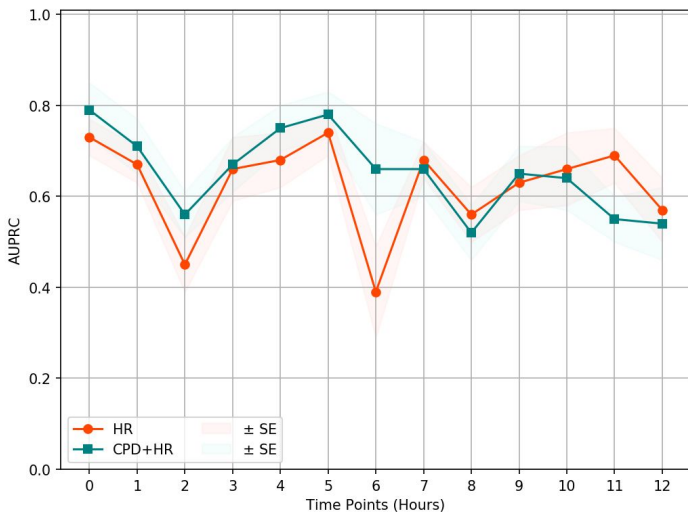
Multiple person detection using YOLOv3 got an IoU of 0.71206. The patient is barely left alone when he/she is uncovered under long hours of observations. To remove any chance of confusion between the temperatures and the body parts that the following segmentation algorithm might face, YOLOv3^[18] was deployed for multiple person detection which achieved an IoU of 0.71206. The other person and the background was then masked to keep the patient in focus for CPD extraction.

Extended data Table 1: Performance of models in the preprocessing steps in the tasks of classification, object detection and segmentation

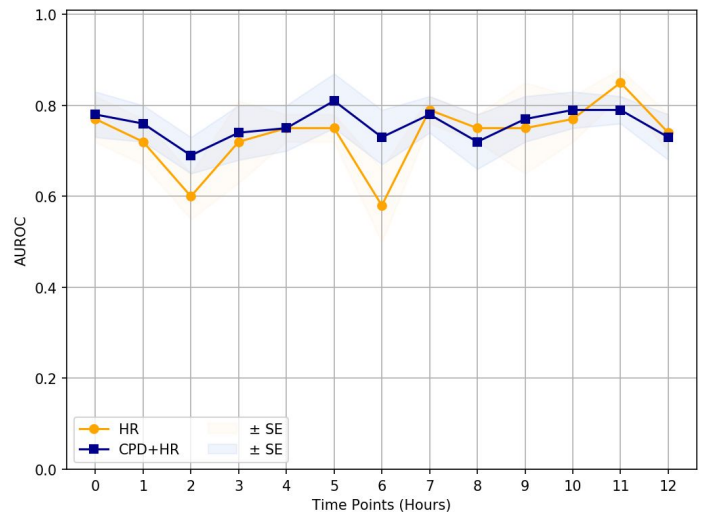
Classification				Object Detection	Segmentation		
ResNet-152				YOLOv3	ResUNet		
Precision	Recall	F1-Score	Accuracy	IoU	BCE Loss	Dice Loss	Total Loss
0.9722	0.9459	0.9589	0.9552	0.7121	0.0692	0.0391	0.0542

Extended data Table 2: Performance of the proposed model predicting the presence of Shock/Non-shock using automated HR only. (S/NS - Number of Shock/Non-Shock sequences present, AUPRC - Area Under Precision Recall Curve, AUROC - Area Under Receiver Operating Characteristics, PPV - Positive predictive value, NPV - Negative predictive value, D - Detection, P - Prediction)

Time Pt.	S/NS	AUPRC	AUROC	Accuracy	Sensitivity	Specificity	PPV	NPV	Youden
		Mean (SE)							Mean
0hr (D)	132, 274	0.73 (0.04)	0.77 (0.05)	0.88 (0.02)	0.75 (0.07)	0.96 (0.01)	0.95 (0.01)	0.86 (0.03)	0.54
1hr (P)	115, 271	0.67 (0.04)	0.72 (0.05)	0.89 (0.02)	0.78 (0.05)	0.95 (0.01)	0.93 (0.01)	0.87 (0.03)	0.53
2hr (P)	125, 253	0.45 (0.06)	0.60 (0.05)	0.84 (0.02)	0.59 (0.07)	0.97 (0.01)	0.92 (0.05)	0.81 (0.02)	0.66
3hr (P)	133, 232	0.66 (0.07)	0.72 (0.09)	0.82 (0.04)	0.69 (0.10)	0.91 (0.04)	0.86 (0.06)	0.84 (0.04)	0.63
4hr (P)	123, 242	0.68 (0.06)	0.75 (0.03)	0.86 (0.04)	0.79 (0.05)	0.93 (0.05)	0.91 (0.06)	0.84 (0.03)	0.45
5hr (P)	120, 247	0.74 (0.05)	0.75 (0.03)	0.91 (0.02)	0.85 (0.04)	0.95 (0.02)	0.92 (0.04)	0.91 (0.02)	0.43
6hr (P)	124, 228	0.39 (0.10)	0.58 (0.08)	0.88 (0.04)	0.79 (0.10)	0.92 (0.04)	0.77 (0.09)	0.93 (0.04)	0.57
7hr (P)	116, 211	0.68 (0.04)	0.79 (0.03)	0.86 (0.03)	0.90 (0.02)	0.82 (0.05)	0.81 (0.04)	0.91 (0.03)	0.56
8hr (P)	85, 216	0.56 (0.06)	0.75 (0.03)	0.88 (0.03)	0.88 (0.05)	0.87 (0.05)	0.85 (0.05)	0.93 (0.02)	0.55
9hr (P)	90, 211	0.63 (0.06)	0.75 (0.10)	0.82 (0.05)	0.78 (0.09)	0.85 (0.06)	0.79 (0.05)	0.86 (0.04)	0.50
10hr (P)	77, 215	0.66 (0.08)	0.77 (0.05)	0.82 (0.05)	0.83 (0.03)	0.82 (0.07)	0.77 (0.08)	0.89 (0.02)	0.62
11hr (P)	62, 226	0.69 (0.06)	0.85 (0.03)	0.92 (0.03)	0.93 (0.03)	0.91 (0.05)	0.86 (0.05)	0.96 (0.01)	0.51
12hr (P)	78, 205	0.57 (0.07)	0.74 (0.04)	0.85 (0.04)	0.75 (0.06)	0.89 (0.04)	0.79 (0.06)	0.89 (0.03)	0.50

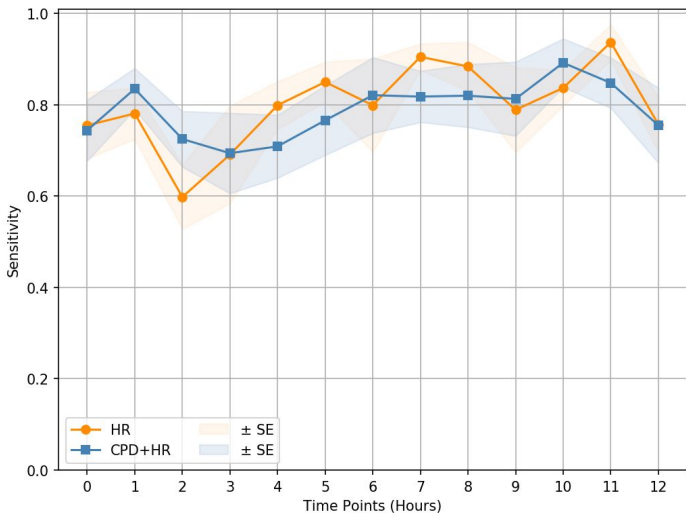


(a) Comparison of AUPRCs

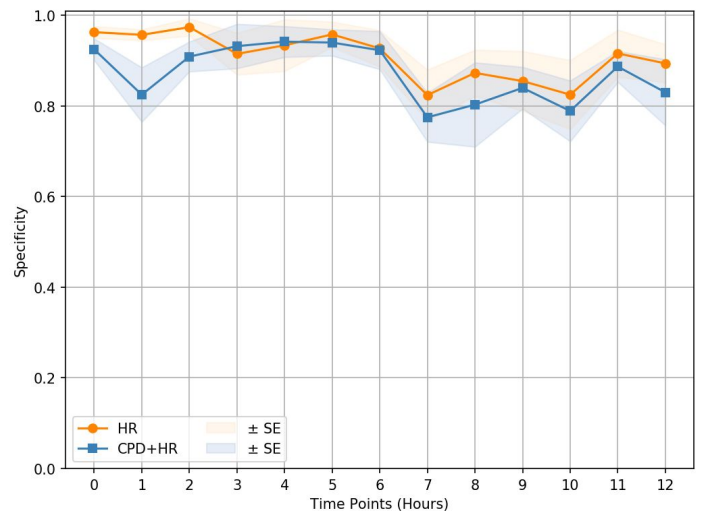


(b) Comparison of AUROCs

Extended data Figure 1: Quality Assessment at Intermediate Time Points. Quality evaluation of the LSTM time series classification models using the AUPRC and AUROC for all 0-12hr predictions.

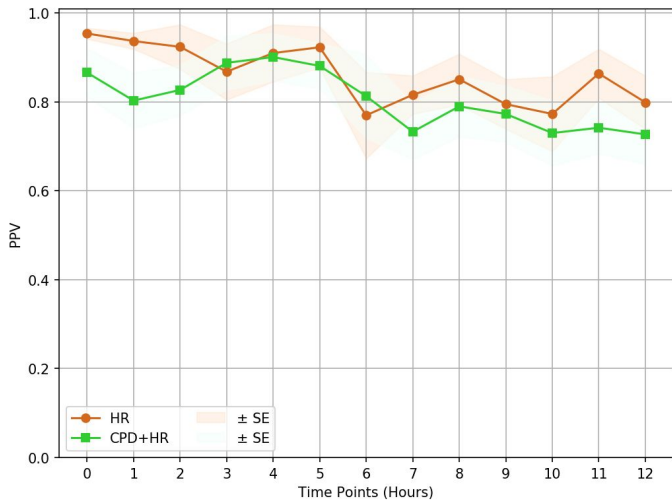


(a) Sensitivity plot

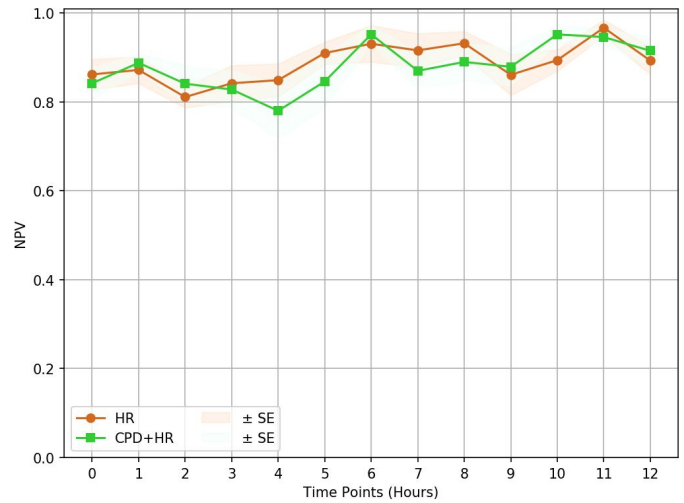


(b) Specificity plot

Extended data Figure 2: Quality Assessment: Sensitivity and Specificity. Comparison of CPD+HR vs HR at each time point till 12 hr. Sensitivity would be a better measure to represent our findings as it takes the False Negatives in consideration. CPD+HR shows a higher trend of sensitivity upto 6 hours.



(a) PPV plot

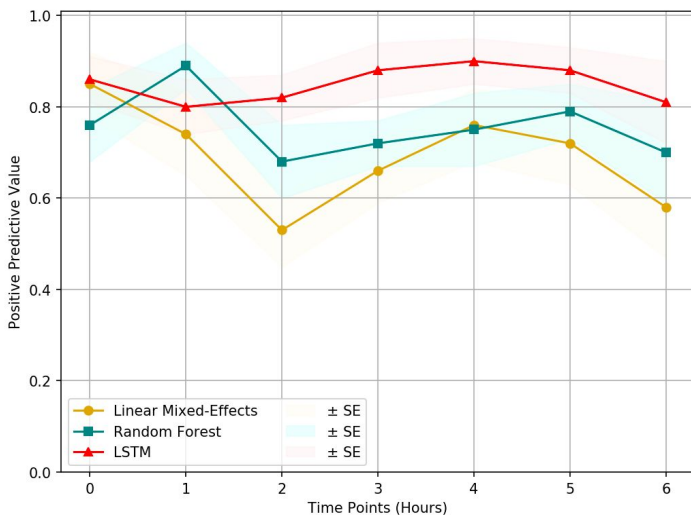


(b) NPV plot

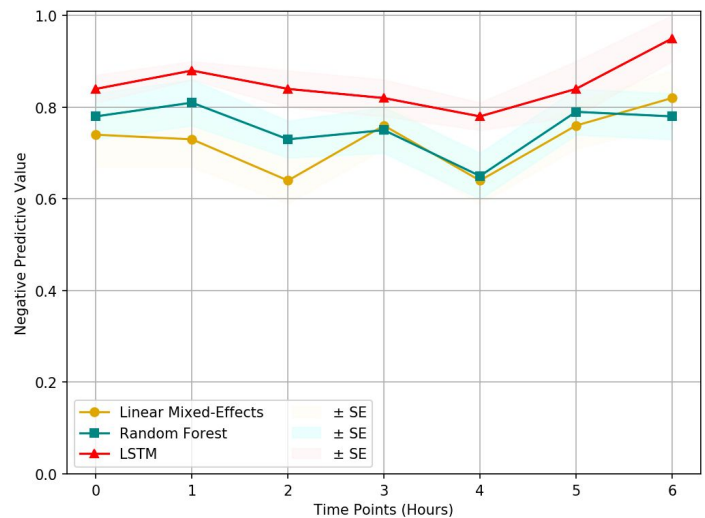
Extended data Figure 3: Quality Assessment: PPV and NPV. Comparison of CPD+HR vs HR at each time point till 12 hr

Extended data Table 3: Statistical significance analysis using the Hanley and McNeil formula^[25]. It is specifically used as our models are stratified k-fold cross validated, and the train and test sets would be independent within each fold, but might not be across the folds. Stratified cross validation violates the IID condition and hence wilcoxon/student's t-test cannot be used. Hanley, et. al. also state that single tailed tests can be used in such cases, and a value near to the cutoff p-value can be considered as an evidence that this observation may not be completely random. Hence, single-tailed analysis was performed with alpha cutoff value of 0.10. It can be observed that AUPRC, our primary metric in use for data imbalance conditions, displays its statistical significance for most of the time points for the time window under our consideration.

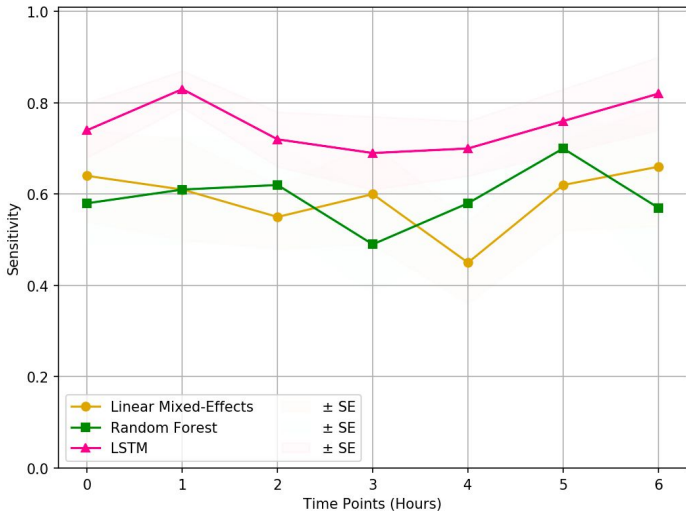
Time Pt.	AUPRC		AUROC	
	p-value	Significant?	p-value	Significant?
0hr	0.116102	~Yes	0.406949	No
1hr	0.174508	~Yes	0.169814	~Yes
2hr	0.002146	Yes	0.006782	Yes
3hr	0.407027	No	0.418395	No
4hr	0.077194	Yes	0.412780	No
5hr	0.215846	No	0.144732	~Yes
6hr	0.000095	Yes	0.014258	Yes



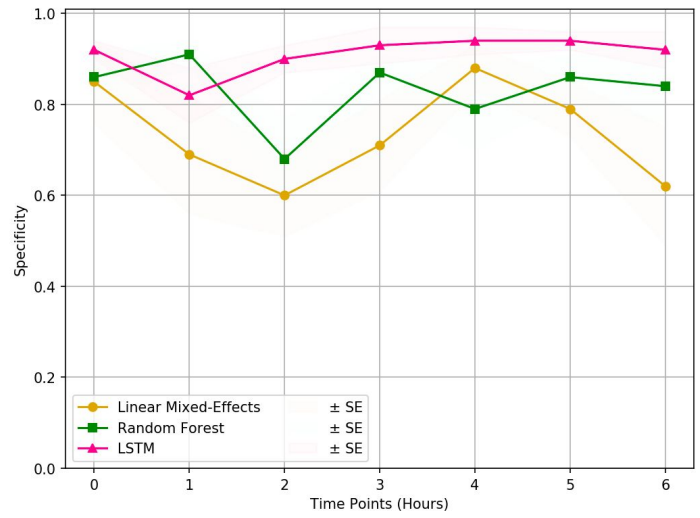
(a) PPV plot



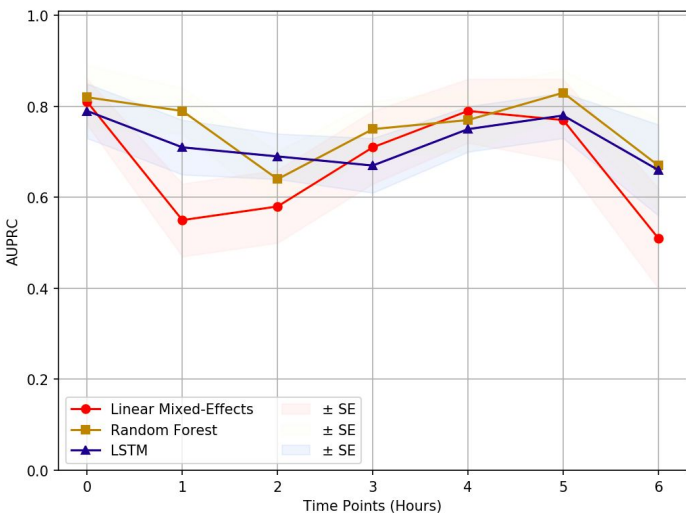
(b) NPV plot



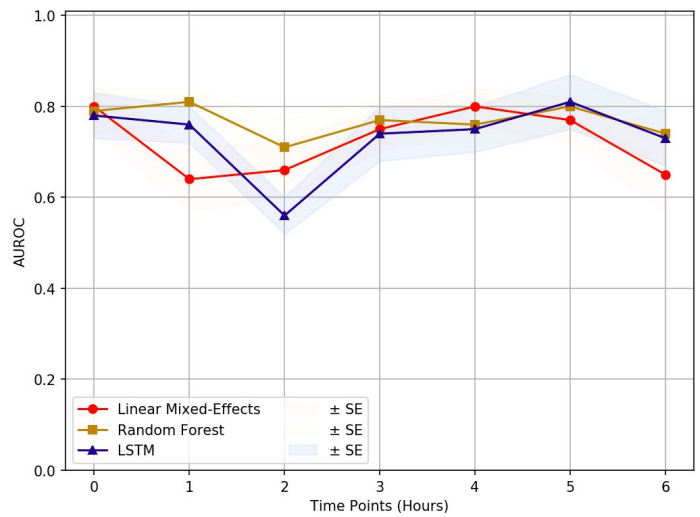
(c) Sensitivity Plot



(d) Specificity Plot



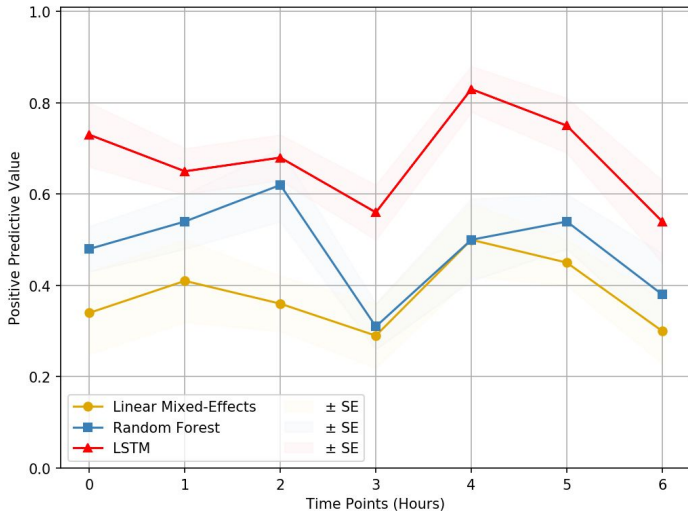
(e) AUPRC plot



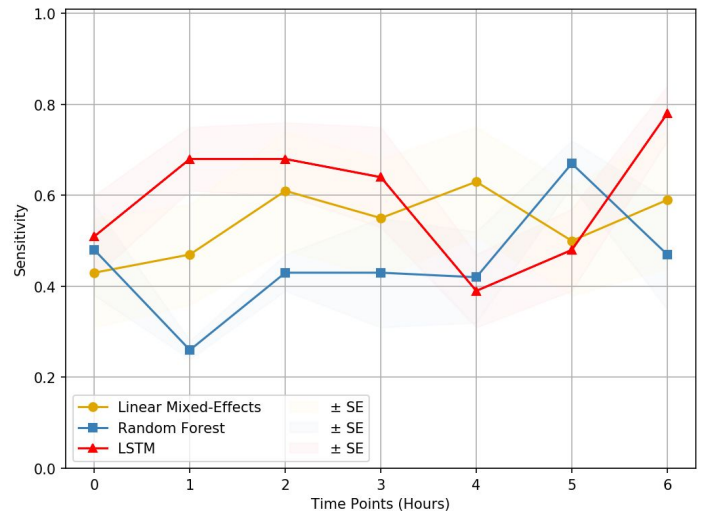
(f) AUROC plot

Extended data Figure 4: Comparison of models for CPD+HR. The three models were compared based on their ability to classify the sequence data. It can be observed that the LSTM model outperforms the other two with respect to the (a) Positive Predictive Value, (b) Negative Predictive Value, (c) Sensitivity and (d) Specificity. (e) Area Under Precision Recall Curve (AUPRC) and (f) Area Under Receiver Operating Characteristics (AUROC) of the models are comparable. Since, our research data exhibits a high data imbalance, F1 score and AUPRC metrics are

the primary considerations. LSTM model displays surpassing F1 score along with PPV, NPV, sensitivity, and specificity, making it the prime choice for this study.



(a) PPV plot



(b) Sensitivity plot

Extended data Figure 5: Comparison of models for CPD only. The three models were compared based on their ability to classify the CPD only sequence data. It can be observed that the LSTM model outperforms the other two in (a) Positive Predictive Value and (b) Sensitivity at most time points.

Extended data Table 3: The metrics observed for direct classification of images for shock detection. (SGD - Stochastic Gradient Descent, AUROC - Area Under Receiver Operating Characteristics)

Model	Optimizer	AUROC	Accuracy
ResNet-50	SGD	0.60	65%

Supplementary Methods S1:

Standard operating protocol (SOP) for capturing thermal videos

1. Before entering the ICU, make sure that one has to wear shoe covers, scrub and a disposable gown.

2. Need to connect the Seek Thermal Camera to the available Smartphone and then open the Seek Thermal app.
3. Ensure that the Seek Thermal app is set to a high-resolution setting and to a “Iron Theme” before capturing videos. Wall mount hanger is used to hold a smartphone. Seek Thermal Camera is placed at a good distance from the bed and at a certain height in-order to make sure that patients on bed are not disturbed.
4. For each patient recorded in this session, following steps were followed:
 - Making the patient comfortable is the most important part. Camera was placed properly and at a good distance from the patient, so that there was no direct contact involved nor any change in patient routine care.
 - In the presence of a nurse, a smartphone is hanged using a wall mount hanger.
 - MSEXcel sheets are used for data collection. Data such as UHID, bed number, date and time of the patients on this session were recorded for the identification purpose.
 - After capturing videos remove the mobile hanger safely and save the data on a server, daily basis, into an identified folder with patient’s UHID, bed information and date.