

Article

The Medical Outcomes Distribution and the Interpretation of Clinical Data Based on C4.5 Algorithm for the RCC Patients in Taiwan

Hsi-Chieh Lee ¹, Ju-Hsia Liu ² and Ching-Sung Ho ^{3,*}

¹ Department of Computer Science and Information Engineering, National Quemoy University, 1 University RD, Jinning Township, Kinmen 89250, Taiwan; imhlee@gmail.com

² Department of Nursing, Jen-Ai Hospital, Taichung 41256, Taiwan; u100077894@cmu.edu.tw

³ Department of Long Term Care, National Quemoy University, 1 University RD, Jinning Township, Kinmen 89250, Taiwan

* Correspondence: csho@nqu.edu.tw; Tel.: +886-8-231-3916

Abstract: The aim of our study is to explore the medical outcomes among patients in the respiratory care center (RCC) and related factors. A cross-sectional study was performed at a regional hospital in central Taiwan from January 2018 to December 2018. The sample consisted of 236 patients who received RCC medical services. The chi-square test, multiple ordinal logistic regression analyses, and C4.5 decision tree algorithm were performed. The risk factors for medical outcomes in critical or deceased patients were obesity (BMI ≥ 27.0) (OR = 2.426, 95% C.I. = 1.106–5.318, $p = 0.027$), being imported from home (OR = 2.104, 95% C.I. = 1.106–3.523, $p = 0.005$), and with the Acute Physiology and Chronic Health Evaluation II (APACHE II) score ≥ 25 (OR = 2.640, 95% C.I. = 1.283–5.433, $p = 0.008$). The results of the C4.5 algorithm showed a precision of 79.80%, a recall of 78.80%, an F-measure of 78.20%, a receiver operating characteristic curve (ROC) area of 89.20%, and a precision-recall curve (PRC) area of 81.70%. It is important to design effective intervention strategies for patients who are obese and with high APACHE II scores and propose timely treatments for the patients' onset of disease at home. Moreover, by using the C4.5 algorithm, data can be interpreted in terms of decision trees to aid the understanding of the medical outcomes of the RCC patients.

Keywords: respiratory care center; machine learning; data mining; C4.5 algorithm; BMI; APACHE II scores; weaning; healthcare; Taiwan medical data



Citation: Lee, H.-C.; Liu, J.-H.; Ho, C.-S. The Medical Outcomes Distribution and the Interpretation of Clinical Data Based on C4.5 Algorithm for the RCC Patients in Taiwan. *Appl. Sci.* **2021**, *11*, 2566. <https://doi.org/10.3390/app11062566>

Academic Editor: Keun Ho Ryu

Received: 25 February 2021

Accepted: 8 March 2021

Published: 12 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Patients ventilated for a prolonged period often consume a high amount of intensive care unit (ICU) resources [1]. It is estimated that the number of ICU patients needing prolonged mechanical ventilation ranges from 3% to 10%, with these patients utilizing 37–40% of ICU resources [2]. Taiwan's National Health Insurance Administration proposed a four-step Integrated Delivery System (IDS) for the care of ICU patients who receive ventilator treatment for longer than 21 days and require lengthy mechanical ventilation. Under this system, patients are transferred from ICUs to respiratory care centers (RCCs). Patients may also be transferred to an Respiratory Care Ward (RCW) if they stay at an RCC for more than 42 days. Respiratory care wards are chronic care units designed for patients with prolonged respiratory failure [3,4]. Due to limited medical resources and the rising cost of health care, disease prognosis has become a very important issue for health sciences.

The Acute Physiology and Chronic Health Evaluation II (APACHE II) score is a severity-of-illness score that is taken at the time of admission of the patient to the ICU, showing the worst figures during the patient's first 24 h of ICU stay [5]. It has been found that there is a significant association between APACHE II scores and risk of mortality (ROM), i.e., the higher the APACHE II score, the higher the ROM [6]. Clarification of the

relationship between APACHE II scores and the medical outcomes of RCC patients can assist in designing more effective medical protocols.

The mechanism by which obesity increases the ROM for ICU patients is unclear. Akinussi et al. [7], for instance, reported that obesity (body mass index (BMI) ≥ 30 kg/m²) was not related to crude ICU mortality when compared with patients with a BMI lower than 30 kg/m². Conversely, Oliveros et al. [8] found lesser mortality for obese patients compared with those with normal weight, but Hogue et al. [9] stated that obesity is not associated with increased risk for ICU mortality. O'Brien et al. [10] demonstrated that lower BMI was linked with higher ROM, whereas overweight and obesity were associated with lower mortality. Additionally, the relationship between BMI and patients in the RCC patients weaning rates is much understudied.

Well-timed access to appropriate care is imperative for critically ill patients. It has, for example, been found that transported patients have a significantly longer ICU stay and greater estimated ROM compared to non-transported patients [11]. The early transfer of patients has also been found to mitigate the risk of adverse impacts on medical outcomes [12]. The transported source of patients may influence the delivery process when referred to the ICU and thus determine the medical outcome.

An exploration of the sociodemographic factors and successful weaning rates' relationship has shown that gender does not affect weaning rates in the Taiwanese context [4]. Frengley et al. [13] showed that the successful weaning rates decrease with patients' age. However, other studies have reported that weaning rates have no relationship with age [14]. The critical issues that affect RCC patients' weaning rates have been shown to be the related causes of respiratory failure as well as confounding diseases such as pulmonary disease, cardiovascular disease, cancer, and renal failure [3].

Data mining algorithms have been developed to collect clinical data to aid in the diagnosis of diseases, such as coronary heart disease, Type 2 diabetes, and Parkinson's disease [15–17]. This is a new scientific area that integrates research from the fields of statistics, machine learning, and computer science (particularly database management). These algorithms facilitate the early detection of diseases, more precise medical treatment, and better use of social medical resources. The C4.5 algorithm is an improved ID3 (Iterative Dichotomiser 3) algorithm, which, in turn, uses information gain rate to select attributes, which produces a more laconic decision tree and enhances the algorithm's efficiency. It has been used in many ways in the fields of medical and clinical informatics. For example, it has been applied to test a clinical guideline-based decision support system in which the C4.5 algorithm was successfully implemented for the testing of a medical DSS relating to chronic diseases [18]. Moreover, it has been adopted for the development of application tools that diagnose diabetes infection symptoms [19].

The aim of our study is to explore the medical outcome distribution among patients who stayed in the RCC with different APACHE II scores and sociodemographic features and to obtain knowledge regarding factors related to the prognosis of RCC patients, especially concerning the roles and effects of BMI and patient sources. The C4.5 decision tree algorithm was performed for the inferential statistical analysis. The performance parameters for the C4.5 decision tree algorithm were precision, recall, F-measure (the harmonic mean of precision and recall), receiver operating characteristic curve (ROC) area, and precision-recall curve (PRC) area.

2. Materials and Methods

2.1. Study Design

A cross-sectional design was employed to achieve the study's research objectives. Clinical data were collected from 236 patients (135 men; 101 women) who received RCC medical services and were transferred from the ICU to the RCC at a regional hospital in central Taiwan between January 2018 and December 2018. Anonymous analysis of the data was used to ensure confidentiality, and the protocol was approved by Jen-Ai Hospital's Medical Ethics Committee (IRB no. 108-74).

2.2. Measurements

Outcome variables were classified according to patients' medical outcomes—critical or deceased, transferred to an RCW, liberated from the ventilator (weaning). The independent variables were obtained from patients' medical records, i.e., age, gender, body weight, and height. The BMI value was calculated using weight and height data in accordance with Taiwan's Health Promotion Administration's BMI classification (<18.5 underweight; 18.5–24.9 healthy; 25–26.9 overweight; ≥ 27 obese). Other factors included the source of admission (home or long-term care facility) and the APACHE II score (divided into categories: <15, 15–24, and ≥ 25). These were employed to analyze the relationship with the medical outcomes of the patient.

2.3. Statistical Analysis

Frequent analyses, mean value, and standard deviation were conducted to assess the distribution of patients' medical outcomes, sociodemographic factors, and APACHE II score. Furthermore, to evaluate the relationship between patients' medical outcomes and related variables, a chi-square test and ordinal logistic regression analysis were performed for the inferential statistical analysis. The SPSS software (SPSS Inc., Chicago, IL, USA) (version 18.0) was used.

C4.5 Decision Tree Algorithm and Performance Evaluation

The C4.5 decision tree algorithm was performed to analyze the data inferentially. Two models were built with five basic attributes plus 12 disease attributes. The performance evaluation parameters for the C4.5 decision tree algorithm were precision, recall, F-measure, ROC area, and PRC area. The C4.5, originally developed by Quinlan [20] and considered an improvement of Quinlan's earlier ID3 algorithm, is used to generate decision trees. The decision trees generated by C4.5 are used mainly for classification. As a result, C4.5 is often referred to as a statistical classifier and a landmark decision tree program that is probably the machine learning workhorse most widely used in practice to date [21]. The C4.5 builds decision trees from a set of training data in a similar way as the ID3, using the concept of information entropy. The training data are a set of pre-classified samples. At each node of the tree, C4.5 picks the data attribute that best splits its sample sets into subsets enriched in one class or the other using the splitting criterion normalized information gain where the highest gain is chosen as the decision criterion. Quinlan [20] suggests using the following equation to calculate the gain ratio:

$$\text{GainRatio}(X, T) = \frac{\text{Gain}(X, T)}{\text{SplitInfo}(X, T)}$$

Considering the information content of a message that indicates—not the class to which the case belongs—but rather the outcome of the test on feature X , we used the following equation for Split Info:

$$\text{SplitInfo}(X, T) = \sum_i^n \left(\frac{|T_i|}{|T|} \log_2 \frac{|T_i|}{|T|} \right)$$

The $\text{GainRatio}(X, T)$ is thus the proportion of information generated by the split that is useful for the classification.

The performance evaluators used for the C4.5 decision tree algorithm were precision, recall, F-measure, ROC area, and PRC area. These are defined below:

Precision (positive predictive value (PPV)): $\frac{TP}{TP + FP}$;

Recall (also known as sensitivity, or true positive rate (TPR)): $\frac{TP}{TP + FN}$;

F-measure (also known as F1 score, which is the harmonic mean of precision and recall): $\frac{2 \times \text{PPV} \times \text{TPR}}{\text{PPV} + \text{TPR}}$;

ROC area: A ROC area is an area under the ROC curve (AUC), one of the common evaluators for machine learning algorithms. A ROC curve is a plot of the false positive rate (x -axis) versus the true positive rate (y -axis) for a number of different candidate threshold values between 0 and 1;

PRC area: A PRC area is the area under the PRC curve, another common performance evaluator for machine learning methods. A PRC curve is a plot of the precision (y -axis) and the recall (x -axis) for different thresholds similar to the ROC curve.

3. Results

3.1. The Medical Outcomes and Related Factors

3.1.1. The Description of the Sample Data

The study included 236 patients (135 men; 101 women). Overall, 139 (58.9%) of the patients were above 75 years old ($M = 74.1 \pm 14.6$ years old). A total of 66 (28.0%) of the patients' negative outcomes were due to cardiac disease, and 24 (10.2%) due to cancer. Moreover, 111 (47.2%) of the patients' BMI values fell in a normal range (18.5–23.9), and 32.3% were obese (≥ 27.0). More patients were transferred from home than from a long-term care facility (54.7%/47.3%). Half (50%) of the samples' APACHE II scores fell in the 15–24 range. Overall, 47.9% of the patients were liberated from the ventilator, 26.3% were admitted to an RCW, and 25.8% of the samples were certified as either critical or deceased. The samples' basic sociodemographic and clinical features are shown in Table 1.

Table 1. Characters and medical results of patients in a respiratory care center (RCC).

	<i>n</i>	%
Gender		
Female	101	42.8
Male	135	57.2
Age	mean \pm SD = 74.1 \pm 14.6	
≤ 64	50	21.2
65–74	47	19.9
75–84	78	33.1
≥ 85	61	25.8
BMI		
Under weight (<18.5)	49	20.9
Normal weight (18.5–23.9)	111	47.2
Over weight (24.0–26.9)	41	17.4
Obese (≥ 27.0)	34	14.5
Source of admission		
Home	129	54.7
Long-term care institution	107	45.3
APACHE II scores	mean \pm SD = 24.58 \pm 7.70	
<15	64	27.8
15–24	115	50.0
≥ 25	51	22.2
Main Diagnosis		
Cancer	24	10.2
Cardiac disease	66	28.0
Pneumonia	41	17.4
Cerebrovascular accident	61	25.8
Diabetes	98	41.5
Lower respiratory illnesses	25	10.6
Hypertension	152	64.4
CKD	42	17.8
Liver disease	15	6.4
Dementia	26	11.0
Parkinson's disease	13	5.5
Miscellaneous	24	10.2
Medical Outcomes		
Critical or deceased	61	25.8
RCW	62	26.3
Weaning	113	47.9
Total	236	100.0

3.1.2. The Inferential Statistical Outcomes

The differences in the patients' medical outcomes distribution resulting from the χ^2 tests are illustrated in Table 2. The results of the univariate analysis showed that the variables that reached statistically significant difference in the distribution of medical outcomes were BMI level ($p = 0.006$), source of admission ($p = 0.011$), APACHE II scores classification ($p = 0.028$), and diabetes ($p = 0.023$). When comparing the differences in medical outcomes, a higher percentage of successful weaning was found in patients with a normal BMI level (51.4%), those being transported from a long-term care institution (56.1%), and those with APACHE II score of less than 15 (56.3%). Conversely, 47.1% of the patients with BMI > 30 and 33.3% of the patients received from home resulted in death or were deemed critical. A higher APACHE II score also resulted in a lower percentage of weaning. No statistically significant differences between groups were found regarding gender, age, and other diseases (Table 2).

Table 2. The distribution of medical results according to different factors.

	Medical Outcome				p Value
	n (%)				
	Critical or Deceased	RCW	Weaning	Total	
Gender					0.486
Female	30 (29.7)	26 (25.7)	45 (44.6)	101 (100)	
Male	31 (23.0)	36 (26.7)	68 (50.4)	135 (100)	
Age					0.102
≤64	10 (20.0)	11 (22.0)	29 (58.0)	50 (100)	
65–74	13 (27.7)	6 (12.8)	28 (59.6)	47 (100)	
75–84	21 (26.9)	26 (33.3)	31 (39.7)	78 (100)	
≥85	17 (27.9)	19 (31.1)	25 (41.0)	61 (100)	
BMI *					0.006
Under weight	13 (26.5)	12 (24.5)	24 (49.0)	49 (100)	
Normal weight	17 (15.3)	37 (33.3)	57 (51.4)	111 (100)	
Over weight	15 (36.6)	8 (19.5)	18 (43.9)	41 (100)	
Fatty	16 (47.1)	5 (14.7)	13 (38.2)	34 (100)	
Source of admission *					0.011
Home	43 (33.3)	33 (25.6)	53 (41.1)	129 (100)	
Long-term care institution	18 (16.8)	29 (27.1)	60 (56.1)	107 (100)	
APACHE II scores *					0.028
<15	11 (17.2)	17 (26.6)	36 (56.3)	64 (100)	
15–24	24 (20.9)	33 (28.7)	58 (50.4)	115 (100)	
≥25	21 (41.2)	12 (23.5)	18 (35.3)	51 (100)	
Cancer					0.364
Yes	9 (37.5)	6 (25.0)	9 (37.5)	24 (100)	
No	52 (24.5)	56 (26.4)	104 (49.1)	212 (100)	
Cardiac disease					0.158
Yes	20 (30.3)	21 (31.8)	25 (37.9)	66 (100)	
No	41 (24.1)	41 (24.1)	88 (51.8)	170 (100)	
Pneumonia					0.512
Yes	8 (19.5)	13 (31.7)	20 (48.8)	41 (100)	
No	53 (27.2)	49 (25.1)	93 (47.7)	195 (100)	
CVA					0.812
Yes	14 (23.0)	16 (26.2)	31 (50.8)	61 (100)	
No	47 (26.9)	46 (26.3)	82 (46.9)	175 (100)	
Diabetes *					0.023
Yes	28 (28.6)	33 (33.7)	37 (37.8)	98 (100)	
No	33 (23.9)	29 (21.0)	76 (55.1)	138 (100)	
Lower respiratory illnesses					0.375
Yes	4 (16.0)	6 (24.0)	15 (60.0)	25 (100)	
No	57 (27.0)	56 (26.5)	98 (46.4)	211 (100)	
Hypertension					0.051
Yes	45 (29.6)	43 (28.3)	64 (42.1)	152 (100)	
No	16 (19.0)	19 (22.6)	49 (58.3)	84 (100)	
CKD					0.211
Yes	14 (33.3)	13 (31.0)	15 (35.7)	42 (100)	
No	47 (24.2)	49 (25.3)	98 (50.5)	194 (100)	
Liver disease					0.171

Table 2. Cont.

	Medical Outcome				p Value
	n (%)				
	Critical or Deceased	RCW	Weaning	Total	
Yes	4 (26.7)	1 (6.7)	10 (66.7)	15 (100)	0.591
No	57 (25.8)	61 (27.6)	103 (46.6)	221 (100)	
Dementia					0.088
Yes	6 (23.1)	9 (34.6)	11 (42.3)	26 (100)	
No	55 (26.2)	53 (25.2)	102 (48.6)	210 (100)	
Parkinson's disease					0.364
Yes	0 (0)	5 (38.5)	8 (61.5)	13 (100)	
No	61 (27.4)	57 (25.6)	105 (47.7)	223 (100)	
Miscellaneous					
Yes	9 (37.5)	6 (25.0)	9 (37.5)	24 (100)	
No	52 (24.5)	56 (26.4)	104 (49.1)	212 (100)	

*: $p < 0.05$.

Table 3 shows the results from the ordinal logistic regression analysis. The results illustrate that BMI, source of admission, and APACHE II scores can be considered as predictors for medical outcomes. Obese patients were 2.426 times (95% C.I. = 1.106 to 5.318) more likely than normal-weight patients to be critical or deceased ($p = 0.027$). Patients who were transported from home were 2.104 times (95% C.I. = 1.257 to 3.523) more so than those from the long-term care institution ($p = 0.005$), and patients with APACHE II scores ≥ 25 were 2.640 times (95% C.I. = 1.283 to 5.433) were similarly found to be so ($p = 0.008$) compared with patients whose APACHE II score < 15 .

Table 3. The ordinal logistic regression analysis of medical outcomes of patients in the RCC.

	OR	95% C.I.	p Value
Medical Outcomes			
Critical or deceased	Reference		
RCW	8.092	4.092–16.000	<0.001
Weaning	2.200	1.171–4.133	0.014
Diabetes			
No	Reference		
Yes	0.622	0.372–1.044	0.072
BMI			
Normal weight	Reference		
Under weight	0.779	0.372–1.630	0.507
Over weight	1.198	0.587–2.445	0.620
Obese *	2.426	1.106–5.318	0.027
Source of admission			
Home **	2.104	1.257–3.523	0.005
Long-term care institution	Reference		
Apache II scores			
<15	Reference		
15–24	0.291	0.758–2.524	0.291
≥ 25 **	2.640	1.283–5.433	0.008
			$R^2 = 0.116$

*: $p < 0.05$, **: $p < 0.01$.

3.1.3. C4.5 Decision Tree Algorithm

Two different models were built using different attributes as the input for the C4.5 algorithm (i.e., the patients' medical data), and the experimental results were derived using 10-fold cross validation.

For Model I, five basic attributes were selected based on the statistical analysis. These included age, APACHE II score, move status, gender, and BMI value. Table 4 and Figure 1 show the performance of the C4.5 decision tree algorithm using these five basic attributes. The results included the following: precision (74.90%), recall (70.80%), F-measure (69.60%), ROC area (85.10%), and PRC area (76%). The decision tree reflects inferential statistical outcomes illustrated above very well. It additionally reveals that younger patients tend to fall into a class with a high probability of successfully weaning.

Table 4. C4.5 algorithm performance for Model I with five basic attributes.

Class	Precision	Recall	F-Measure	ROC Area	PRC Area
Weaning	64.40%	91.20%	75.50%	83.70%	80.40%
Critical Death	81.40%	57.40%	67.30%	88.80%	74.00%
RCW	87.90%	46.80%	61.10%	84.00%	70.10%
Weighted Avg.	74.90%	70.80%	69.60%	85.10%	76.00%

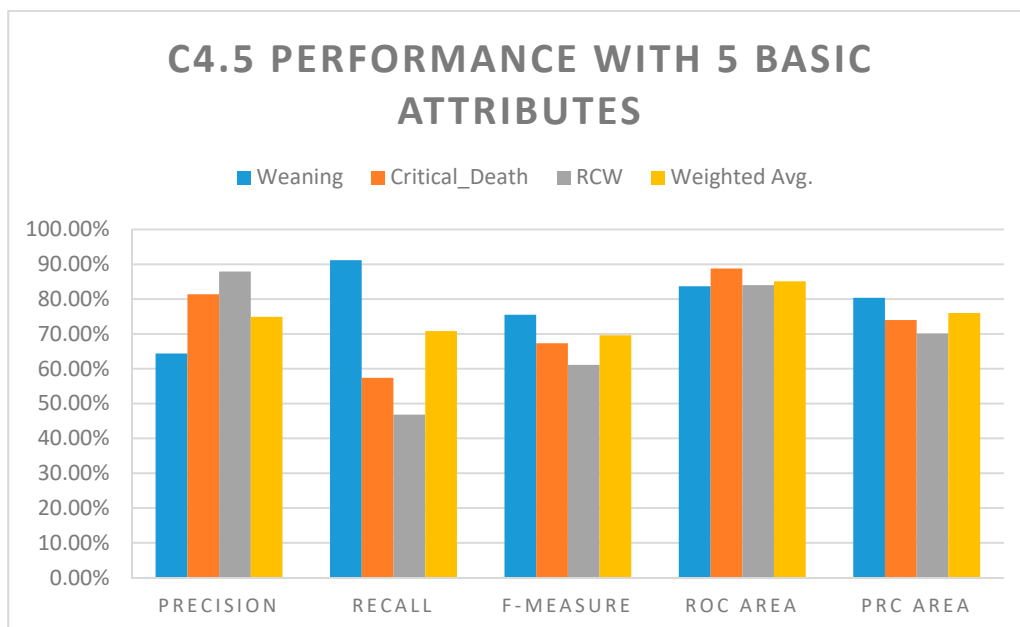
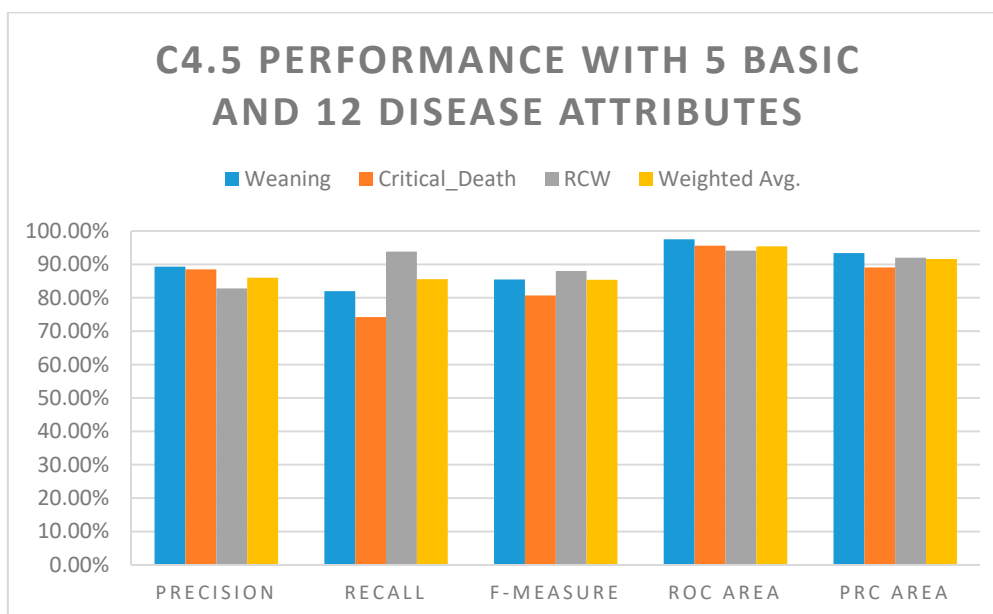


Figure 1. C4.5 algorithm performance with five basic attributes.

For Model II, in addition to the five basic attributes adopted in Model I, 12 disease-related factors were added as input for the C4.5 decision tree algorithm. These included cancer, cardiac disease, pneumonia, cerebrovascular accident, diabetes, lower respiratory illnesses, hypertension, chronic kidney disease (CKD), liver disease, dementia, Parkinson’s disease, and other chronic diseases. Table 5 and Figure 2 show the performance of the C4.5 decision tree algorithm with its 5 basic and 12 disease attributes. The results of the C4.5 algorithm improved as follows: precision (79.80%), recall (78.80%), F-measure (78.20%), ROC area (89.20%), and PRC area (81.70%).

Table 5. C4.5 algorithm performance for Model II with 5 basic and 12 disease attributes.

Class	Precision	Recall	F-Measure	ROC Area	PRC Area
Weaning	75.50%	92.90%	83.30%	88.70%	83.60%
Critical Death	86.00%	60.70%	71.20%	88.90%	78.90%
RCW	81.50%	71.00%	75.90%	90.60%	81.10%
Weighted Avg.	79.80%	78.80%	78.20%	89.20%	81.70%

**Figure 2.** C4.5 algorithm performance for Model II with 5 basic and 12 disease attributes.

4. Discussion

Patients ventilated for prolonged periods often experience progressive chronic respiratory failure and substantial comorbidity. They also consume a large percentage of intensive care resources. It is, therefore, necessary to identify the significant factors related to RCC patients' medical outcomes. The results of this study indicated that BMI, APACHE II scores, and patient source are related to the medical outcomes of RCC patients. Compared to patients with successful weaning, critical or deceased patients were characterized by being obese (BMI value ≥ 27.0), transferred from home, and having APACHE II scores ≥ 25 .

In ventilated patients, obesity (BMI value ≥ 27.0) was found to be related to higher odds of death or critical medical outcomes compared with normal-weight patients, whereas overweight (BMI value between 24.0 and 26.9) and underweight patients (BMI value < 18.5) were not found to be significantly associated with the distribution of medical outcomes. This result implies that obesity may play an important role in the prognosis of RCC patients. These results are related to the extant literature. Mancuso [22], for instance, suggested that obesity may be a significant factor for the pathogenesis of pulmonary diseases, which contains pro-inflammatory mediators produced in the adipose tissue that induces a state of systemic inflammation [22]. Some researchers have also stated the probability of a higher frequency of organ dysfunctions in obese patients [23–25]. However, other studies have found that a BMI less than 18.5 kg/m^2 was associated with increased mortality in ICU patients [26,27]. The possible reasons for this may be that BMI status was presented to be an indicator of general nutrition, and sufficient nutrition is important to help respiratory muscle contractility and the successive ability to expectorate phlegm. As such, the effect of obesity related to medical outcomes of RCC patients needs further investigation.

The APACHE II scoring system was developed by Knaus et al. in 1985 [5]. It was calculated by employing the values of 12 physiological variables, age, and chronic health status, each of which was assigned points using a similar tactic. Scores were allotted to the worst values of each variable for the calculation of the APACHE II scoring system. Studies have demonstrated that a growing score is closely associated with ROM in ICU patients and the outcome in a wide range of disease conditions [6,28,29]. The results of our study similarly demonstrate a significant association between APACHE II scores and the medical outcome of RCC patients. It was found that the higher the APACHE II score, the higher the risk of being critical or deceased. These findings confirm the proficiency of this scoring system to predict patients' outcomes according to the degree of severity of their disease.

Individuals living at home are usually thought to be healthier and to have a better ability to take care of themselves than those staying in long-term care institutions. However, the results of this study demonstrate that RCC patients who were transported from home had worse medical outcomes than patients who were transferred from long-term care facilities. The risk of becoming critical or ROM for transported from home patients was found to be more than twofold (OR = 2.135). This may be attributable to a delay in receiving medical treatment incurred by patients transported from home. This result is also supported by the extant literature in that it has been found that transportation barriers may play an important role, e.g., creating delayed access to health care may cause extended suffering, tough and exorbitant treatment, and increased morbidity and mortality whereas intervention strategies such as the early detection of diseases and timely treatments can save patients transported from home [30]. Therefore, the reasons behind the poor medical outcomes of RCC patients referred from home require further investigation.

The results of the C4.5 decision tree algorithm suggest that by adding the 12 disease attributes to the analysis, the precision improved almost 5% from 74.90% to 79.80%. The recall rate also improved significantly, from 70.80% to 78.80%. The other three performance measurements, namely, F-measure, ROC area, and PRC area, also improved, as shown in Table 6 and Figure 3. These data can assist medical personnel in recognizing the workings of the system for the prognosis of RCC patients. As such, medical personnel can recognize the result and intervene in the structure of the resulting decision tree. Future studies can also be conducted to assist with the interpretation of patient data in order to facilitate high rates of true positive and negative results.

Schönhofer, et al. (2002) found age to be associated with increased mortality in ventilated patients [31], i.e., ventilator-dependent patients older than 80 usually experience extremely poor medical outcomes. Our study, however, showed that age is not a risk factor related to the medical result of RCC patients; consequently, higher age was not definitely found to be associated with worse health statuses. As such, there may be other factors related to the medical outcomes of RCC patients, and thus, more research is needed to explore the effect of age on the prognosis of RCC patients.

As with all research, our study has limitations. One is that the study employed a cross-sectional design. As such, the effect of factors related to RCC patients' medical outcomes needs further examination in that the study sample originated from one metropolitan–regional hospital in central Taiwan, and thus, the findings may not be generalizable to RCC patients throughout Taiwan. Another weakness is the small case number of some categories used to evaluate the relationship between the variables. Despite these weaknesses, our study contributes to the literature in that it explored the relationship between the medical outcomes' distribution and health issues such as APACHE II scores and BMI status levels, as admission sources for RCC patients in Taiwan.

Table 6. C4.5 algorithm performance comparison for Model I and Model II.

Model	Precision	Recall	F-Measure	ROC Area	PRC Area
Model I (5 basic attributes)	74.90%	70.80%	69.60%	85.10%	76.00%
Model II (5 basic + 12 disease attributes)	79.80%	78.80%	78.20%	89.20%	81.70%

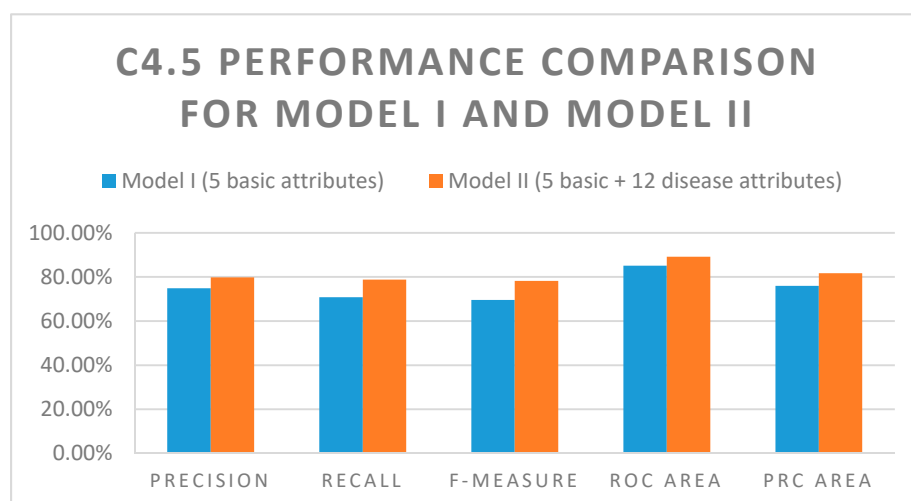


Figure 3. C4.5 algorithm performance comparison for Model I and Model II.

5. Conclusions

Weaning from RCC suggests that the patients tolerate spontaneous breathing. Since patients ventilated for a prolonged period usually consume a high amount of medical resources, our study explored and identified important factors related to the medical outcomes and using C4.5 algorithm data mining technology in which clinical data can be interpreted in terms of a decision tree to aid in the understanding of the medical outcomes of the RCC patients. We also suggest that new intervention strategies should be designed to care for patients who are obese and have high APACHE II scores and that more effective methods for timely treatments for the patients' onset of the disease at home.

Author Contributions: Conceptualization, H.-C.L. and C.-S.H.; methodology, H.-C.L. and C.-S.H.; formal analysis, H.-C.L. and C.-S.H.; investigation, J.-H.L. and C.-S.H.; writing—original draft preparation, H.-C.L. and C.-S.H.; writing—review and editing, H.-C.L., J.-H.L. and C.-S.H.; project administration, J.-H.L. and C.-S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study protocol was approved by the Medical Ethics Committee of Jen-Ai Hospital (IRB no. 108-74).

Informed Consent Statement: The institutional review board of the Medical Ethics Committee of Jen-Ai Hospital permitted this study without the requirement of written informed approval from any of the study patients.

Data Availability Statement: The data are available by contacting the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bach, P.B.; Carson, S.S.; Leff, A. Outcomes and resource utilization for patients with prolonged critical illness managed by university-based or community-based subspecialists. *Am. J. Respir. Crit. Care Med.* **1998**, *158*, 1410–1415. [[CrossRef](#)]
- Wu, Y.; Kao, K.; Hsu, K.; Hsieh, M.; Tsai, Y. Predictors of successful weaning from prolonged mechanical ventilation in Taiwan. *Respir. Med.* **2009**, *103*, 1189. [[CrossRef](#)]
- Su, J.; Lin, C.-Y.; Chen, P.-J.; Lin, F.J.; Chen, S.-K.; Kuo, H.-T. Experience with a step-down respiratory care center at a tertiary referral medical center in Taiwan. *J. Crit. Care* **2006**, *21*, 156–161. [[CrossRef](#)]
- Yang, P.-H.; Hung, J.-Y.; Yang, C.-J.; Tsai, J.-R.; Wang, T.-H.; Lee, J.-C.; Huang, M.-S. Successful weaning predictors in a respiratory care center in Taiwan. *Kaohsiung J. Med. Sci.* **2008**, *24*, 85–91. [[CrossRef](#)]
- Knaus, W.A.; Draper, E.A.; Wagner, D.P.; Zimmerman, J.E. APACHE II: A severity of disease classification system. *Crit. Care Med.* **1985**, *13*, 818–829. [[CrossRef](#)]
- Naved, S.A.; Siddiqui, S.; Khan, F.H. APACHE-II score correlation with mortality and length of stay in an intensive care unit. *J. Coll. Physicians Surg. Pak.* **2011**, *21*, 4.

7. Akinnusi, M.E.; Pineda, L.A.; El Solh, A.A. Effect of obesity on intensive care morbidity and mortality: A meta-analysis. *Crit. Care Med.* **2008**, *36*, 151–158. [[CrossRef](#)]
8. Oliveros, H.; Villamor, E. Obesity and mortality in critically ill adults: A systematic review and meta-analysis. *Obesity* **2008**, *16*, 515–521. [[CrossRef](#)] [[PubMed](#)]
9. Hogue, C.W.; Stearns, J.D.; Colantuoni, E.; Robinson, K.A.; Stierer, T.; Mitter, N.; Pronovost, P.J.; Needham, D.M. The impact of obesity on outcomes after critical illness: A meta-analysis. *Intensive Care Med.* **2009**, *35*, 1152. [[CrossRef](#)] [[PubMed](#)]
10. O'Brien, J.M.; Welsh, C.H.; Fish, R.H.; Ancukiewicz, M.; Kramer, A.M. Excess body weight is not independently associated with outcome in mechanically ventilated patients with acute lung injury. *Ann. Intern. Med.* **2004**, *140*, 338–345. [[CrossRef](#)] [[PubMed](#)]
11. Flabouris, A. Patient referral and transportation to a regional tertiary ICU: Patient demographics, severity of illness and outcome comparison with non-transported patients. *Anaesth. Intensive Care* **1999**, *27*, 385. [[CrossRef](#)] [[PubMed](#)]
12. Gerber, D.R.; Schorr, C.; Ahmed, I.; Dellinger, R.P.; Parrillo, J. Location of patients before transfer to a tertiary care intensive care unit: Impact on outcome. *J. Crit. Care* **2009**, *24*, 108–113. [[CrossRef](#)] [[PubMed](#)]
13. Dermot Frengley, J.; Sansone, G.R.; Shakya, K.; Kaner, R.J. Prolonged mechanical ventilation in 540 seriously ill older adults: Effects of increasing age on clinical outcomes and survival. *J. Am. Geriatr. Soc.* **2014**, *62*, 1–9. [[CrossRef](#)] [[PubMed](#)]
14. Huang, C. How prolonged mechanical ventilation is a neglected disease in chest medicine: A study of prolonged mechanical ventilation based on 6 years of experience in Taiwan. *Ther. Adv. Respir. Dis.* **2019**, *13*, 1753466619878552. [[CrossRef](#)] [[PubMed](#)]
15. Lamy, J.-B.; Ellini, A.; Ebrahimi, V.; Zucker, J.-D.; Falcoff, H.; Venot, A. Use of the C4.5 machine learning algorithm to test a clinical guideline-based decision support system. *Stud. Health Technol. Inform.* **2008**, *136*, 223.
16. Khaleel, A.H.; Al-Suhai, G.A.; Hussan, B.M. Application Tool based on C4.5 Decision Tree for Diagnosing Diabetes Infection Symptoms. *J. Commun. Technol. Electron. Comput. Sci.* **2019**, *22*, 7–15.
17. Ramani, R.G.; Sivagami, G. Parkinson disease classification using data mining algorithms. *Int. J. Comput. Appl.* **2011**, *32*, 17–22.
18. Sa'Di, S.; Maleki, A.; Hashemi, R.; Panbechi, Z.; Chalabi, K. Comparison of data mining algorithms in the diagnosis of type II diabetes. *Int. J. Comput. Sci. Appl. (IJCSA)* **2015**, *5*, 1–12. [[CrossRef](#)]
19. Wiharto, W.; Kusnanto, H.; Herianto, H. Interpretation of clinical data based on C4.5 algorithm for the diagnosis of coronary heart disease. *Healthc. Inform. Res.* **2016**, *22*, 186–195. [[CrossRef](#)] [[PubMed](#)]
20. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers: Boston, MA, USA, 1993.
21. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann: San Francisco, CA, USA, 2011; p. 191.
22. Mancuso, P. Obesity and lung inflammation. *J. Appl. Physiol.* **2010**, *108*, 722–728. [[CrossRef](#)] [[PubMed](#)]
23. Brown, C.V.; Neville, A.L.; Rhee, P.; Salim, A.; Velmahos, G.C.; Demetriades, D. The impact of obesity on the outcomes of 1153 critically injured blunt trauma patients. *J. Trauma Acute Care Surg.* **2005**, *59*, 1048–1051. [[CrossRef](#)]
24. Ciesla, D.J.; Moore, E.E.; Johnson, J.L.; Burch, J.M.; Cothren, C.C.; Sauaia, A. Obesity increases risk of organ failure after severe trauma. *J. Am. Coll. Surg.* **2006**, *203*, 539–545. [[CrossRef](#)] [[PubMed](#)]
25. Neville, A.L.; Brown, C.V.; Weng, J.; Demetriades, D.; Velmahos, G.C. Obesity is an independent risk factor of mortality in severely injured blunt trauma patients. *Arch. Surg.* **2004**, *139*, 983–987. [[CrossRef](#)]
26. Gupta, R.; Knobel, D.; Gunabushanam, V.; Agaba, E.; Ritter, G.; Marini, C.; Barrera, R. The effect of low body mass index on outcome in critically ill surgical patients. *Nutr. Clin. Pract.* **2011**, *26*, 593–597. [[CrossRef](#)] [[PubMed](#)]
27. Pickkers, P.; de Keizer, N.; Dusseljee, J.; Weerheijm, D.; van der Hoeven, J.G.; Peek, N. Body mass index is associated with hospital mortality in critically ill patients: An observational cohort study. *Crit. Care Med.* **2013**, *41*, 1878–1883. [[CrossRef](#)] [[PubMed](#)]
28. Donnino, M.W.; Saliccioli, J.D.; Dejam, A.; Giberson, T.; Giberson, B.; Cristia, C.; Gautam, S.; Cocchi, M.N. APACHE II scoring to predict outcome in post-cardiac arrest. *Resuscitation* **2013**, *84*, 651–656. [[CrossRef](#)]
29. Johnson, C.; Toh, S.; Campbell, M. Combination of APACHE-II score and an obesity score (APACHE-O) for the prediction of severe acute pancreatitis. *Pancreatology* **2004**, *4*, 1–6. [[CrossRef](#)] [[PubMed](#)]
30. Syed, S.T.; Gerber, B.S.; Sharp, L.K. Traveling towards disease: Transportation barriers to health care access. *J. Community Health* **2013**, *38*, 976–993. [[CrossRef](#)]
31. Schönhofer, B.; Euteneuer, S.; Nava, S.; Suchi, S.; Köhler, D. Survival of mechanically ventilated patients admitted to a specialised weaning centre. *Intensive Care Med.* **2002**, *28*, 908–916. [[CrossRef](#)]