Chapter 4

DISCUSSION AND CONCLUSION

4.1 Discussion

Specific Entries result and General Entries result are both using same 31 patients. The only difference is that the latter use all 7460 rows of data including the one that does not have in the journal entries.

Two models has been developed for both LLSM and LDCM for each case where the first one use zero value in place of missing entries (case A) and the other one use mean value for each features (case B). The reason I use two models was because the missing entries gave an undetermined status or noise for the patient's data. Therefore, by replacing them with some value it becomes possible to estimate the performance.

Figure 3.1 and 3.2 shows the model of LLSM for case A and case B. The outcome of the models show that they are not much difference in term of the model itself and the features selected by the feature selection algorithm. The model has an axis of Status Risk obtained from journal entries (SRJ) and an axis of Status Risk obtained from DSS (SRD). They are plotted in the same figure to see how they relate to each other. Status Risk has been classified into 3 classes which are class 1, class 2 and class 3. Class 1 is set for patients who are admitted into the hospital or inpatient. Meanwhile class 2 is set for patients who are visited by general practitioner (GP) out of schedule time or they themselves went to GP. Whereas class 3 is for any change in medicine or any worsen condition reported. Ideally, we want a result like class 1 should be

tabulated heavily in class 1 region for each axis. Same thing should happen for class 2 and class 3. Clearly, we can see for both models, most of the data tabulated between class 1 and class 2 in SRD axis for class 1 in SRJ axis. Same goes for class 2 in SRJ axis. This gives us quite a disappointment outcome. The data tabulated heavily between 2 classes and it seems difficult for us to exactly determine which classes they are belong to. Class 3 gives us an unquestionable bad result. Since the models show about the same thing, we can expect the feature selection also be the same for both. It just differs in the priority of selection as show in table 3.1.

Figure 3.3 and 3.4 gives us an expected bad result since we include all the data which are not in the journal entries. It is just to see how the plot would be look like. 1 extra class is created in this plot, which is class 0. Class 0 means that the data are included in the unknown status due to missing data. Since the missing entries are too many it affect other result from class 1, 2 and 3. We can omit this result since they not provide a good impression at all.

Table 3.3 and 3.4 show the confusion table for specific entries result for both case A and case B. Ideally, we want the all the tabulation only in the diagonal direction of the table as shown in table 4.1. But in this case, we can't see the pattern here as the tabulation spread all over the place. The same thing happened for general entries result case; in fact it is much worse as illustrated in table 3.5 and 3.6.

The performance of the system (Cohen's Kappa Coefficient, κ) is illustrated in table 3.9. The coefficients value of less than 0.1 shows that the system gives only slight agreement between rater A and rater B (see Appendix A). Two cases have the coefficient of 0.008 which clearly tell us that this system have no agreement at all. This outcome was expected due to several reasons.

	Status Risk (Journal Entries)			
Status Risk (DSS)	**1**	**2**	**3**	**Total**
1	*150 (0.478)*	6 (0.019)	0 (0.000)	156 (0.497)
2	3 (0.010)	*120 (0.382)*	1 (0.003)	124 (0.395)
3	3 (0.010)	3 (0.010)	*28 (0.089)*	34 (0.108)
Total	156 (0.497)	129 (0.411)	29 (0.092)	*314 (1.000)*

Table 4.1: Example of ideal confusion table

4.2 Problems

There are a few problems have been identified which lead to the displeasing result in this study. The details will be discussed below.

1. The main problem of this study is the journal entries data. The journal does have records about the patient's condition and some other notes. But these entries don't have the standard so that one might not be able to assess the patient's conditions easily using some algorithm. In other words, these entries can only be evaluated manually rows by rows and then translated into language that is recognized by the computer. This will result of a very inefficient way to deal with the data.
2. The journal contains many missing entries that are represented by 'null' value. The missing entries greatly affect the analysis since we can't really determine a way to deal with this 'null' value. This was demonstrated in this study where we replace the 'null' value with zero and mean of each feature in two different cases.

4.3 Recommendation

This study is just a beginning to the actual big project. Since we know the initial problems, we can try to improve this study by altering some of the methods or more importantly improving the input. In our case, the input is the journals entries. Below is the summarize recommendation that might be implemented to help improving this study.

1. A feature might be added into the journal entries that would summarize the status of the patient's instead of depending only on the memo or notes provided. This can give an overview of the patient's health status and at the same time might be a great help to develop an algorithm to classify the patient's into groups or classes as we did in this study.
2. Since we only use data from trends records in this study, we might miss out important features. Thus, by providing some extra features from other records on a particular patient's, we might get a better result.
3. Missing entries represented by 'null' value is quite a big problem for us at the moment and we can't find a good way to deal with it yet. Probably we can use mean value of 2 weeks data for each features instead of mean value for the whole set of data. This way, we might get some trends if the missing entries are constant for several days.

4.4 Conclusion

The main concern of this study is to see the reliability of the data provided by the journal entries. This is just a beginning phase of the study for further development to improve the decision support system currently in use. The result shows there is only slight agreement between DSS and journal entries data. This means that the reliability of the data can be questionable. The outcome of this study was expected as there are so many missing entries in the journal. In addition, the patient's status in the journal entries is not really clear and must be interpreted by someone capable. However, this study can be improved by several ways as discussed in previous section. An example is by introducing some extra features instead of just the trends of data as we used in this study. It is hoped that the patient's health can be monitored more effectively by improving this study and consequently upgrading the current decision support system.

Appendix A

COHEN'S KAPPA COEFFICIENT

Cohen's Kappa Coefficient (κ) is a statistical measure of inter-rater agreement. In this report, the performance of the system was analyzed using this method where the two raters in my case are the score result from DSS and the outcome of risk status from journal entries. κ takes into account the agreement occurring by chance and therefore is thought to be a more robust measure than usual simple percentage agreement. The value of κ is varies from 0 to 1 and the following table is given by Landis and Koch for interpreting κ values [5].

κ	**Interpretation**
< 0	No agreement
0.0 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Table A.1: Interpretation of κ value

The equation for κ is:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \qquad (A.1)$$

$\Pr(a)$ is the total proportion of observer agreement, $\Pr(e)$ is the probability that agreement is due to chance.

Pr (a) is given by

$$\Pr(a) = \sum_{i=1}^{m} p_{ii} \qquad (A.2)$$

Meanwhile the Pr (e) is given by

$$\Pr(e) = \sum_{i=1}^{m} p_{i.}p_{.i} \qquad (A.3)$$

Assuming we have two raters, A and B which have can have n observation and can be classify into one of m classes. Table A.1 below shows the confusion table that can be used to calculate κ value. n_{ij} means that the number of times Rater A classified an observation as class i and at the same time Rater B classified as class j. p_{ij} in the parenthesis is an alternative way to represent this value which is the fraction of total n and can be obtained by dividing n_{ij} by n [6].

	Rater B				
Rater A	A	B	...	*m*	**Total**
A	$n_{11}(p_{11})$	$n_{12}(p_{12})$	...	$n_{1m}(p_{1m})$	$n_{1.}(p_{1.})$
B	$n_{21}(p_{11})$	$n_{22}(p_{22})$	...	$n_{2m}(p_{1m})$	$n_{2.}(p_{2.})$
:	:	:	…	:	:
m	$n_{m1}(p_{m1})$	$n_{m2}(p_{m2})$	...	$n_{mm}(p_{mm})$	$n_{m.}(p_{m.})$
Total	$n_{.1}(p_{.1})$	$n_{.2}(p_{.2})$	...	$n_{.m}(p_{.m})$	$n(1)$

Table A.2: Confusion table for calculating Cohen's Kappa Coefficient

	Status Risk (Journal Entries)			
Status Risk (DSS)	**A**	**B**	**C**	**Total**
A	115 (0.366)	85 (0.271)	15 (0.048)	215 (0.685)
B	30 (0.096)	32 (0.102)	8 (0.025)	70 (0.223)
C	11 (0.035)	12 (0.038)	6 (0.019)	29 (0.092)
Total	156 (0.497)	129 (0.411)	29 (0.092)	314 (1)

Table A.3: Example of confusion matrix for calculating κ

Using equation A.1, A.2 and A.3, we can now calculate the value of κ.

$$\Pr(a) = 0.366 + 0.102 + 0.019 = 0.487$$

$$\Pr(e) = (0.685)(0.497) + (0.223)(0.411) + (0.092)(0.092) = 0.441$$

$$\kappa = \frac{0.487 - 0.441}{1 - 0.441} = 0.082$$

REFERENCES

1. Wyatt, J. and D. Spiegelhalter, *Field trials of medical decision-aids: potential problems and solutions.* Proceedings - the Annual Symposium on Computer Applications in Medical Care, 1991: p. 3-7.
2. Delaney, B.C., et al., *Can computerised decision support systems deliver improved quality in primary care?* BMJ: British Medical Journal, 1999. **319**(7220): p. 1281.
3. M.S. Srivastava, E.M.C., *Applied Multivariate Statistics.* North Holland Amsterdam, 1983.
4. Pudil, P., J. Novovicová, and J. Kittler, *Floating search methods in feature selection.* Pattern Recognition Letters, 1994. **15**(11): p. 1119-1125.
5. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**(1): p. 159-174.
6. Redmond, S.J. and C. Heneghan, *Cardiorespiratory-Based Sleep Staging in Subjects With Obstructive Sleep Apnea.* Biomedical Engineering, IEEE Transactions on, 2006. **53**(3): p. 485-496.